Universal Access in the Information Society (2023) 22:1–35
https://doi.org/10.1007/s10209-021-00823-1

LONG PAPER

# Machine translation from text to sign language: a systematic review

Navroz Kaur Kahlon[1] · Williamjeet Singh[1]

Accepted: 23 June 2021 / Published online: 3 July 2021


**Abstract**

An equal opportunity for all is the basic right of every human being. The deaf society of the world needs to have access to all the information just like hearing people do. For this to happen there should be a mode of direct communication between hearing and deaf people. The need at this time is to automate this communication so as the deaf society is not dependent upon human interpreters. This paper deals with the systematic survey of conventional and state-of-the-art sign language machine translation and sign language generation projects. We used a standard procedure of carrying out a systematic literature review on 148 studies published in 30 reputed journals and 40 premium conferences and workshops. Existing literature about sign language machine translation is broadly classified into three different categories. These categories are further sub-classified into different classifications depending upon the type of machine translation. Studies pertaining to the specified classifications have been presented with their advantages and limitations. Different methods for sign language generation are reported with their benefits and limitations. Manual and automatic evaluation methods used in every study is presented along with their respective performance metrics. We call for increased efforts in presentation of signs to make them an easy and comfortable mode of communication for deaf society. There is also a requirement to improve translation methods and include the contribution of advanced technologies such as deep learning and neural networks to make an optimal translation process between text and sign language.

**Keywords** Sign language · Deaf people communication · Machine translation · Systematic literature review · Sign language generation

## 1 Introduction and motivation

From the dawn of time, communication has played a significant role in the evolution of our species. Today one cannot fathom carrying on day-to-day businesses and life without a language understood by all the parties involved. Those with a hearing disability have developed sign languages to overcome the spoken language barrier, but still face many daily life challenges. Two deaf people with no other physical disability can communicate with each other using sign language, but if a hearing person has to communicate with a deaf person or vice-versa, the communication has many barriers. In such cases, a translation process is required, which can translate the hearing person's spoken language to the sign language of the deaf person and vice-versa (conversion of sign language to speech). A human translator is an option for language translation, but it is an expensive option and not always available. Therefore, a practical solution is required to fulfil the everyday needs of people.

Machine translation can break this linguistic barrier by automating the translation process. This advancement is essential for Deaf people as the translation provides a platform for communication between deaf and hearing people and provides deaf people with the same opportunities to access information as everyone else [126].

Several research projects in machine translation to sign language have been developed, such as ViSiCAST, a European project that aims to provide improved access to deaf citizens through sign language presented by virtual animation or avatar [2]. Another European project in this league is the eSIGN Project. DePaul University initiated a significant avatar-based project for American Sign Language, which created an avatar named "Paula" to portray all linguistic parameters of ASL while translating English to ASL [3]. Another project ProDeaf, converts Portuguese text and voice

✉ Navroz Kaur Kahlon
kahlon.navroz3@gmail.com

1 Department of Computer Science and Engineering, Punjabi University, Patiala, Punjab 147002, India

to Portuguese Sign Language (LIBRAS) to improve communication between deaf and hearing.

There are various projects like SignAloud, Kinect Sign Language Translator, SignAll, and MotionSavvy, translating the signed words or sentences to spoken language. However, in this paper, we will review only the research papers which deal with text/Voice to sign language translation.

### 1.1 Motivation for work

Following are some factors that worked as the motivation behind this study:

- Sign Language Machine Translation (SLMT) is a fundamental idea to make information accessible for the deaf community an easy task. Our study discusses various conventional and contemporary approaches along with their shortcomings.
- Our study explores various kinds of SLMT methods and the sign synthesis required to perceive the final output for the deaf community. Furthermore, we will discuss multiple evaluation methods and performance metrics used by different studies.
- The lack of a complete systematic literature review in the field of SLMT was a motivating factor. We analyzed the entire existing database for SLMT and summarized it to report certain opportunities for future investigations.

This main goal of this review was to classify the existing literature focusing on different SLMT systems, sign synthesis, and evaluation methods applied in various studies.

## 2 Review method

The systematic review reported in this paper was done following the guidelines discussed in [4–6]. The steps included in the review include developing a review protocol, conducting the study, and discussing the findings and future prospective.

The systematic review shows different classifications and sub-classifications of SLMT interspersed over time.

Following is a table with a set of research questions that were required to plan this review.

### 2.1 Planning the review

The review protocol formulated includes the research question framework, the databases searched, and search strings used to search for relevant studies required for this review. The review method was discussed between both the authors and then finalized after discussion. Electronic databases were extensively searched and the studies extracted are reported. The results were extracted using a search engine of scholarly documents Google Scholar. The search strings required the titles to have the keywords like "sign language," "machine translation," and "sign generation/synthesis." While searching several studies mentioned above, studies related to sign language recognition were excluded as the review conducted only cites the studies based upon text to sign language translation. The lack of systematic studies on SLMT is the primary motivation behind this review.

### 2.2 Research questions

This main goal of this systematic review's was to identify and classify the existing literature focusing on SLMT, sign synthesis, and performance metrics used to evaluate the translation techniques. Planning the review required a set of research questions. Table 1 lists the specific research questions and sub-questions. Every research question is supported with motivation for considering that question.

### 2.3 Sources of information

A broad perspective is required for extensive and broad coverage of literature. Before starting the review, an appropriate set of databases must be chosen to make the findings of highly relevant articles probable. The following electronic databases were searched for this review:

- Springer (https://link.springer.com/)
- ScienceDirect (https://www.sciencedirect.com/)
- IEEE explore (https://ieeexplore.ieee.org/Xplore/home.jsp)
- Taylor and Francis (https://www.tandfonline.com/)
- ACM Digital Library (https://www.acm.org/publications/digital-library)

### 2.4 Search criteria

In nearly all the searches, the keywords "machine translation," "sign language," and "sign generation" are included in the abstract. We tried to extract as much relevant literature as possible from the e-resources mentioned. We undertook a pedantic database search to ensure the comprehensiveness of our review. Table 2 lists the e-resources and the search strings used to extract relevant papers. Many known research papers were not included in the review mainly because our review is one-sided, i.e., from text to sign language and not vice-versa. Only papers in English were included in our study. The studies' abstracts and titles are used as the initial filtration parameter to find out relevant studies.

**Table 1** Research questions, sub-questions and motivation

| Research questions | Motivation |
|---|---|
| 1. What is the conventional status of Sign Language Machine translation? | It helps in understanding the initial research endeavours in the field. Various state-of-art works have been reported with their respective advantages and limitations |
| 2. What is the current status of Rule-Based Machine Translation?<br>2.1 What are the categories of RBMT, what are the methods used for pre-processing, translation, and generation of signs? | The study of various RBMT research papers helps understand the role of linguistic understanding of both source and destination languages. We have mentioned studies that picked upon all three kinds of RBMT and the number of studies that followed rule-based translation strategy |
| 3. Research status of Corpus-Based Machine Translation?<br>3.1 What are different categories of corpus-based machine translation, methods used for pre-processing, translation and sign generation, and bilingual corpora's importance? | The study of corpus-based machine translation research papers answers the need to move from rule-based to corpus-based approaches. The research question explores the studies which have adopted different kinds of data-driven strategies for SLMT |
| 4. What are the studies discussing the current scenario of sign synthesis?<br>4.1 What are various types of modes used for sign synthesis?<br>4.2 What are different types of annotation systems and platforms used for synthesis using a 3D avatar? | It helps in knowing the final output of the translation process. It is essential to know the kind of modes available for sign synthesis. The studies discussed show the advantages of 3D avatars over other types of sign generation outputs |
| 5. What are the evaluation methods used in different studies?<br>5.1 What are the different types of evaluation methods?<br>5.2 What are performance metrics used under each evaluation method? | It is essential to understand the evaluation measures adopted to measure the efficiency of the system. Various performance metrics and the number of studies for each performance metric are also reported |

**Table 2** Search strings in e-resources

| Sr. no. | E-resource | Search string | Dates | Product/content type |
|---|---|---|---|---|
| 1 | ieeexplore.ieee.org | Abstract: "sign language", "machine translation", "sign generation" | All dates | Conferences and Journals |
| 2 | www.acm.org | Abstract: "sign language", "machine translation", "sign generation" | All dates | Journals and Proceedings |
| 3 | https://www.sciencedirect.com/ | Abstract: "sign language", "machine translation", "sign generation" | All dates | All sources |
| 4 | https://link.springer.com/ | Abstract: "sign language", "machine translation", "sign generation" | All dates | All sources |
| 5 | https://www.tandfonline.com/ | Abstract: "sign language", "machine translation", "sign generation | All dates | Journals |

### 2.5 Inclusion and exclusion criteria

In the first stage, irrelevant papers were excluded based on their titles. In our study, machine translation is a broad topic, including translation between spoken languages and from sign to spoken languages. These studies were ineligible for review as we were focussing on text to SLMT. The systematic review included qualitative and quantitative research studies published up to and including 2020. Research papers repeated in different e-resources were individually excluded to remove any redundancy.

### 2.6 Quality assessment

After the first filtering of relevant papers, a quality assessment was performed on the papers. The quality assessment was conducted based upon guidelines mentioned in [83], i.e., each study was assessed for bias, internal and external validity of results.

Using the quality assessment as per "Appendix A", all of the included papers contain premium SLMT research, thus increasing the selected database's validity. In quality assessment after the screening questions of A.1 and A.2 were considered, the study for detailed findings was performed in A.3

## 3 Background

Automatic sign language processing comprises different applications, including sign language translation and sign synthesis [128]. In this section, we will discuss various terminology related to SLMT. We explain sign language, sign language challenges, the taxonomy of SLMT, and summarize the translation process between spoken languages to sign language.

### 3.1 Sign language

Sign languages are full-fledged natural languages with their own grammar and lexicon [135]. They are expressed through manual articulations in combination with non-manual components.

There are two kinds of features to be considered in signs: *manual features* and *non-manual features*. Manual features include the movement of hands, fingers, and arms. Non-manual features hold a fundamental component in all sign languages. They include facial expressions, eye gaze, head, and upper body movement, and position. Non-manual signs, in combination with manual signs, give a complete representation of sign language.

Sign languages like spoken languages organize elementary units called phonemes into meaningful units called semantic units. These units are represented through hand shape, orientation, location, movement, and non-manual expressions. Unlike spoken languages, sign languages take advantage of the spatial nature of the language through the use of classifiers. Classifiers help in spatially showing referent type, size, shape, and movement.

There is a common misconception that sign languages are dependent on spoken sign languages. As sign language develops, it borrows some elements from spoken languages but how much is borrowed varies. The grammar of sign language is different from spoken language. Therefore, while translating from a spoken language to sign language, both the source and target languages' grammar must be considered for an adequate translation.

### 3.2 Machine translation

Machine translation is a computer science field that investigates the use of software to translate text/voice from one language to another. It can assist in breaking linguistic barriers and providing easy access to information. A human translator can be substituted with machine translation software to perform this translation. In [8] the author describes different types of machine translations listed below:

- Rule-Based Machine Translation (RBMT): This translation is based upon linguistic rules, and it involves information about the source and target language. It is further classified into the following categories:
    - Direct-based: This consists of translating word by word using a bilingual dictionary after applying morphological analysis;
    - Interlingua-based: In this, the source text is transformed to an Interlingua (i.e., abstract language independent representation) presentation, and from this presentation, the target text is generated;
    - Transfer-based: This is similar to Interlingua based translation as both have intermediate representation; the difference is that in interlingua-based intermediate representation is independent of the languages in question, whereas, in transfer-based, it has some dependency on the language pair involved.
- Corpus-Based machine translation (CBMT): This type of translation requires bilingual data i.e., data of both spoken and sign languages. It is further classified as follows:
    - Statistical `Machine Translation (SMT): This process uses probability and is based on bilingual text corpora. It is dependent upon large parallel context;
    - Example-Based Machine Translation (EBMT): It is based on the idea of analogy and requires a number of bilingual examples without any linguistic knowledge;
    - Hybrid Machine Translation: This involves the strength of multiple machine translation approaches within a single machine translation system.
- Neural Machine Translation (NMT): This method of translation uses an artificial neural network to predict. It uses an encoder that divides the sentence into constituent words, and the meanings are represented using vectors. The sentences are then interpreted as a whole and are further decoded using the weighted distribution over the encoded vectors. [78].

### 3.3 Challenges of sign language (SL)

Sign languages differ from spoken language in various ways. Apart from the grammatical differences, sign and spoken languages differ in structure, word order, and lexicon for some languages. San-Segundo et al. point specific issues while translating Spanish to Spanish Sign Language (LSE), such as mapping one semantic concept to a specific sign, mapping several semantic concepts onto a unique sign, and generating several signs from one semantic concept [134]. Similarly, in translation from Arabic text to Arabic Sign Language (ArSL), similar challenges can be noticed wherein the difference in grammatical rules and difference in the word order of both source and target languages pose issues in the translation process [102].

Another vital difference between spoken and sign languages is sequencing [95, 164]; in spoken languages, phonemes are produced in sequence, whereas sign languages have non-sequential components because fingers, hands, and face movements can be involved in sign language simultaneously. Similar differences can be found between Thai and Thai sign language (TSL) wherein the Thai language is linear, but Thai sign language is simultaneous with parallel, temporal and spatial configurations [30]. TSL, like other sign languages, differs from the Thai language in word order as a Thai sentence contains Subject (S), Verb (V), and Object (O) in a sequence that differs from TSL.

The above differences specify that sign languages have some specific features, which are listed below [126]:

- Non-Manual components: The articulators in sign languages are hands, arms, face, head, neck, and body. While signing, the signer covers all the non-manual components to perform a sign;
- The use of space: The space around the signer is a lexical space with a phonological value and is used to articulate signs;
- Parts of Speech: The sign languages have different parts of speech, and sometimes the noun, the adjective, and the verb are represented by the same sign;
- Classifiers: Classifiers are used to give meaning to a linguistic category of nouns and nominal's. In sign languages, classifiers are certain hand shapes that substitute for other signs and have morphological value;
- Syntax: Not all languages organize the sentence structure in SVO order; many sign languages have different placement of object and verb in the sentence.

### 3.4 Process of text to sign language machine translation

The automated sign language translation systems make more information and services accessible to deaf and hearing-impaired people in an economical way [30].

In this paper, we will discuss translations from spoken languages to sign languages. Each country generally has its native sign language like American Sign Language (ASL), British Sign Language (BSL), Spanish Sign Language (LSE), Indian Sign Language (ISL), etc. To translate from one language to another, parallel corpora is to be created where corresponding sentences, phrases, or words from the two languages can be identified [123].

Text to sign language process is divided into three modules:

- The first module pre-processes the input text in which the text is morpho-syntactically analysed and broken into words with their types (noun, verb, particle, etc.) with a language model's help;
- In the second module, the pre-processed words are converted into sign sequences. For this module, various machine translation strategies (RBMT, EBMT, SMT, HBMT, and NMT) can be followed;
- The last module converts the sign sequences into sign glosses, videos, or animated avatars.

The following section of the paper discusses the current status of machine translation in sign language. It includes conventional as well as state of the art studies in this field. The following studies are milestones in the field of sign language generation, some having high citations.

## 4 Current status of machine translation in sign language

Several SLMT projects have been carried out throughout the world like TESSA, ViSiCAST, TEAM, ZARDOZ, SASL-MT, TGT [12, 130, 149, 156, 164, 165]. All the projects mentioned above have played a significant role in making machine translation an essential technology text to sign language translation systems. In the next part of the paper, we will discuss different machine translation studies and some contemporary works in this field. Table 3 shows the categorization of machine translation with the count of studies in each category.

**Table 3** Number of studies referring to different categories of machine translation

| Sr. no. | Type of machine translation | Code | # | Citations |
|---|---|---|---|---|
| 1 | Rule Based Machine Translation | C1 | 38 | [12, 53, 66, 76, 87, 129, 130, 149, 152, 156, 161, 164, 165], [9, 51, 54, 62, 64, 104–106, 131, 133, 166, 167], [4, 17, 29, 42, 44, 46, 48, 81, 88, 102, 120, 126] |
| 2 | Example Based Machine Translation | C2 | 5 | [5, 15, 114, 127, 139] |
| 3 | Statistical Machine Translation | C3 | *10* | [3, 22, 23, 30, 92, 95, 110, 112, 146, 162] |
| 4 | Hybrid Machine Translations | C4 | *9* | [2, 19, 79, 89, 97, 113, 115, 117, 132, 162] |
| 5 | Neural Machine Translation | C5 | 6 | [20, 103, 136–138, 147] |

### 4.1 Pre-processing

The language processing model is an essential prerequisite for machine translation from spoken to sign language. A primary language processing model includes input text parser, Word eliminator, stemmer, and phrase reordering modules [6]. A parser applies tokenization to get parts of speech (POS) tagging for the input text. Newly generated tokens are further transferred to the eliminator module, removing all the unwanted tokens from the parsed text. The stemmer module works on verbs and converts every word into a simple present tense of the verb. As mentioned above that spoken and sign language differ from each other in terms of sequencing. This difference gives rise to the phrase reordering model, which reorders the arguments according to sign language grammar [95].

Language models or the pre-processing model reduces variability in the source language necessary in both rule and corpus-based systems. Language models perform morphological analyses on source data to produce accurate and maximum rules and significantly reduce source-target language alignment in corpus-based systems.

The earliest use of language models were seen in systems like ZARDOZ [154, 156], TEAM [164], ASL Workbench [144] and ViSiCAST [13, 104, 129, 130]. The TEAM system employed Synchronous Tree Adjoining Grammar (STAG) [140] rules to build an English dependency tree during the analysis stage. ASL Workbench used Lexical Functional Grammar (LFG) to convert English texts into functional structures, further converted to ASL output. ViSiCAST system, which is also considered the most innovative system amongst those mentioned above, uses CMU Link Parser [141] to analyze English input text. Prolog declarative clause grammar rules are further used to convert the output into a Discourse Representation Structure (DRS) [68].

Earlier parsers and morpho analyzers were based on English syntax. As more and more researchers worldwide started working in this field, several multilingual parsers and specific language parsers were also developed. Parsers like ILSP parser, VnTokenizer, and the Al-Khalil morpho system were explicitly made for Greek, Vietnamese and Arabic language respectively [1, 93, 124]. Stanford Parser is another present-day parser that, apart from English, is adapted to work on other languages like Chinese, German and Arabic [153].

Figure 4 shows some parsers used by various studies of machine translations. Table 3 shows different machine translations referred to in Fig. 4 as CX, where X is a number from 1 to 5. The reference numbers of the corresponding study are also mentioned in Fig. 4. Tables 4, 5, 6, 7, and 8 give a comparative analysis of various studies related to different types of machine translations. Studies in these tables are picked from Table 3, and the sign synthesis modes used by the corresponding works are also presented.

### 4.2 Types of machine translations

During 1990s, the need for a sign language corpus and translating the spoken language into a sign language started surfacing. Some significant works during this time have been done in different languages like ZARDOZ for English to other sign languages, TESSA for British to British sign language, INGIT for English to Indian Sign language, TEAM for English to American sign language, and SYUWAN for Japanese to Japanese sign language [76, 152, 156, 161, 164].

Sign translation raises various exciting issues for how machine translation can be used to make it an automatic translation. Machine translation helps synthesize existing translation technologies into a workable and socially relevant application [156].

Another critical factor in SLMT program is to measure the suitability of the translation program. Several efforts have been made in measuring translation systems' performance, considering both automatic and manual evaluation methods. The most common evaluation metrics which are used in the studies are WER (Word Error Rate), PER (position-independent word error rate), BLEU (Bilingual Evaluation under Study), TER (Translation Error Rate) [121, 125, 142, 151]. Various other performance metrics and criteria of manual evaluation will be discussed in detail in Sect. 4.4.

Different approaches of machine language have come up in past years for text to sign language translation which can be broadly categorized as (1) rule-based machine translation and (2) corpus-based machine translation, which is sub-categorized as (a) example-based machine translation (b) statistical-based machine translation (c) hybrid-based machine translation and, (3) neural machine translation. All approaches have their strengths and weaknesses (Fig. 1).

Translation systems under these approaches will be discussed in the following part of the paper. Figure 2 gives the taxonomy of Machine Translation.

#### 4.2.1 Rule-based machine translation

Rule-based Machine Translation (RBMT) system is based upon linguistic information about the source and target language. Rule-based systems take sentences of the source language as input and generate them to output sentences based on morphological, syntactic, and semantic analysis of both source and target languages. RBMT system consists of a source language morphological analyser, a source language parser and translator, a target language morphological generator, and a target language parser for composing the output sentences. The Vauquois Pyramid presented in Fig. 3 is used to describe the complexity and sophistication of the rule based approaches [87].

Earlier approaches of rule based translation include some outstanding works like the ZARDOZ system, the SL

**Table 4** Projects following rule-based machine translation and their comparative analysis

| Project | Description | Languages | Year | Strengths | Limitations | Sign synthesis |
|---|---|---|---|---|---|---|
| [156] | Uses Interlingua approach for translation | English to Irish, British and Japanese Sign Language | 1994 | Complete infrastructure including parsing, interlingua generation and animation component has been implemented | Comprehensive grammar for all languages has not been developed | Animation sequences |
| [164] | Prototype MT system is made based upon linguistic, spatial and visual information | English to American Sign Language(ASL) | 2000 | Takes into account visual and spatial information associated with ASL signs | No evaluation of the system has been done | Animations using Parallel Transition Networks (Pat-Nets) [8] |
| [161] | Translation on the basis of phrase lookup approach instead of language translation | British to British Sign Language (BSL) | 2000 | Demonstrates virtual signers with high fidelity to deliver legible signing | Issues in language translation and understanding | Animation is achieved by Motion Capture (Simon-the-Signer) |
| [129] | Syntactic, semantic and discourse oriented NLP techniques are used | English to Dutch, German and British Sign Languages | 2000 | Uses word sense disambiguation algorithms to resolve many linguistic issues | No sign synthesis, no evaluation | DRS representation |
| [104] | Techniques involved in extracting information from link grammar and DRS representation are detailed | English to different sign languages | 2001 | Link Dictionary definition produces large scale DRS component | No sign Synthesis, no evaluation | DRS representation |
| [54] | ASCII-Stokoe notation is taken as input which is converted to an intermediate feature tree and then converted to 3D animation in VRML | English to ASL | 2002 | The software developed is free, open, flexible and easy to expand | Rotation angles of certain parts (hands, wrist, and elbow) are difficult to determine | 3D avatar in VRML |
| [43] | Uses Discourse Representation(DRS) structures as intermediate semantic representation | English to BSL | 2002 | Synthetically creates signing character without using motion capture | – | Head Driven Phase Structure Grammar (HPSG) linked with SiGML, |
| [106] | HPSG sign language generation component of the prototype system is described | English to BSL | 2004 | Uses HamNoSys description of signs which further can drive an avatar | No evaluation of the system | 3D avatar |
| [62] | Focuses on classifier Predicates and combines all direct, transfer and Interlingua approaches | English to ASL | 2004 | Illustrates a 3D scene representation | No ASL output for classifier predicates | 3D model based on NLI animation planning process |
| [105] | Preliminary consideration of SL plurals and placement within the SL grammar | English to BSL | 2005 | HPSG formalism assists HamNoSys in generating directional verbs | The system undergoes no evaluation | 3D animation |
| [51] | Multi-modal application which provides user interface to the deaf client | English to BSL | 2006 | Uses synthetic animation and user interface | Facial expressions and lip movements of 3D avatar faced criticism | 3D virtual animation |
| [165] | Follows the translation design of TEAM project | English to South African Sign Language (SASL) | 2006 | Pronoun resolution algorithm places objects in signing space | Lack of interface between machine translation and signing avatar component | H-Anim 3D avatar |

**Table 4** (continued)

| Project | Description | Languages | Year | Strengths | Limitations | Sign synthesis |
|---|---|---|---|---|---|---|
| [166] | Uses STAG parser (earlier used in TEAM) in order to generate non-manual signs | English to SASL | 2006 | Stress patterns are included to improve non-manual generation of signs | Signing space construction needs improvement | 3D avatar |
| [131] | Applied 153 rules for translation and GER and BLEU are used for evaluating results | Spanish to Spanish Sign Language(LSE) | 2006 | Conducts field experiments using automatic performance metrics | Errors in speech recognition inhibit generation of gestures | 3D avatar (VGuido) |
| [133] | Gesture sequence generation and gesture animation are the main focus | Spanish to LSE | 2007 | Develops an animated agent and strategy for reducing gesture design time | AGR agent is not sophisticated | AGR(Agent for Gesture Animation) avatar made up of geometric shapes |
| [46] | Combines natural language knowledge, machine translation and avatar technology for dynamic generation of signs | Greek to Greek Sign Language (GSL) | 2007 | Analysis adopted for GSL gives multilayer information required for performance of grammatical GSL utterance | The system provides optimum result in a restricted sub-language oriented environment | 3D avatar |
| [9] | Uses separate module for analyzing and transformation of input text, also the model includes mood preferences of the deaf person | Spanish to LSE | 2009 | Inclusion of mood of signer in translation | Evaluation conducted is informal | Maxine animation engine |
| [88] | Detailed implementation of language processing component and focuses upon problems of knowledge extraction of SL grammar | Greek to Greek Sign Language (GSL) | 2010 | GSL conversion module is open-source, platform independent and functional through Web | GSL conversion module works with a language specific parser | 3D sign representation |
| [126] | Follows transfer based approach and applies word order generation algorithm to deal with topic oriented surface order of LSE | Spanish to LSE | 2014 | Follows algorithms to cover semantic and lexical gaps during translation | Use of glosses as final output | Glosses |
| [4] | Several information extraction techniques are applied before being passed onto animation stage for sentiment analysis | Portuguese to Portuguese Sign Language | 2014 | Use of open source software for animation | Less linguistic understanding of the language | 3D Avatar using Blender |
| [48] | Formalizes LSF production Rules and triggers them with Text pre-processing | French to French Sign Language (LSF) | 2015 | Rules are formed based upon corpus analysis and not influenced by text structure | Proposed evaluation model is premature | 3D avatar |
| [17] | Web application for sign language generation. Linguistic model is used | French to French Sign Language (LSF) | 2016 | Focuses on both manual and non-manual articulators | Not available for public access | WebGL for 3D animation |
| [42] | Presents implementation of post-processing stage to grammar based machine translation | Greek to GSL | 2016 | Incorporation of SL editor in MT system which increases the accessibility potential of deaf users | Evaluators demanded availability of synonyms in the editing environment | 3D avatar using HamNoSys transcription |

**Table 4** (continued)

| Project | Description | Languages | Year | Strengths | Limitations | Sign synthesis |
|---|---|---|---|---|---|---|
| [87] | A large and good quality of parallel corpora is created for sign language | Greek to GSL | 2018 | Produces good language models which can be used in SMT approaches | Only glosses are generated as final output | Glosses |
| [102] | Semantic system is developed which performs lexical, semantic and syntactic analysis on the input sentence | Arabic to Arabic Sign Language (ArSL) | 2019 | Creation of parallel corpus in health domain which is freely available for researchers | No animation for SL output is produced | Gloss and GIF images |
| [29] | Uses Machine Learning Decision Tree to convert structured sentences into Short SL forms | Vietnamese to Vietnamese Sign Language (VSL) | 2019 | Use of decision tree to shorten complex Vietnamese sentences | Database is small resulting in lesser accuracy, | 3D avatar using HamNoSys transcription |

**Table 5** Projects using example-based machine translation and their comparative analysis

| Project | Description | Language | Year | Strengths | Limitations | Sign Synthesis |
|---|---|---|---|---|---|---|
| [114] | One of the initial approaches in the field of EBMT in SLMT. Use of Marker Hypothesis for segmentation of English text | English to Dutch Sign Language | 2005 | Segmentation approaches used provided similar chunks of data which helped in alignment during translation | Small corpus and dictionary size. No 3D animation for SL output | Annotations using ELAN tool |
| [5] | Morphological analyzer and root extractor are used along with chunk-based EBMT to overcome small corpora problems | Arabic to ArSL | 2011 | Addition of morphological analyzer and root extractor to the system | Small corpus leads to several missing signs error problem | Sign clips |
| [15] | EBMT coupled with Genetic algorithm to produce naturalized animations | Any text to ASL | 2012 | Use of specific form and emotion interpolation factors in SL animation | Automatic management of signing area is not included | 3D humanoid using Sign Modelling Language |
| [139] | Lexical Supervision component is used in learning and translating component of the procedures | Turkish to Turkish Sign Language | 2017 | Use of k-fold cross-validation method for evaluation to determine training and test sets | Use of glosses as final SL output | TSL glosses |
| [127] | Translation happens based on word type to reduce the problem of unexpected words in the sentence | Vietnamese to Vietnamese Sign Language | 2018 | Text processing resulted in better accuracy | Long processing time | VSL glosses |

**Table 6** Projects using Statistical Machine Translation and their comparative analysis

| Project | Description | Language | Year | Strengths | Limitations | Sign Synthesis |
|---|---|---|---|---|---|---|
| [22] | A bilingual corpus trains the translation system, and methods to develop the SL corpora are demonstrated | German to German Sign Language (DGS) | 2004 | Produces satisfactory results on a small corpus | Low translation performance | Glosses using annotation |
| [146] | Morpho-syntactic knowledge source of German is used to improve the translation quality | German to DGS | 2006 | Use of flexible POS parser allowing transforming words based on lexical assumptions | Poorly supported avatar | Glosses<br>Avatar |
| [92] | Follows phrase based translation which uses word reordering to speed up the process | Czech to Czech Sign Language | 2007 | The sign editor provides online access for creating and editing signs | Lack of intelligibility in isolated signs | Animation |
| [30] | Syntax and semantic differences between source and target languages are considered in developing this model | Thai to Thai Sign Language | 2007 | The model is simple, modular, accurate and user-friendly | No animation or 3D avatar as SL output | Pictures with text |
| [162] | Viterbi algorithms are applied along with context free grammars in this three stage translation process | Chinese to Taiwanese Sign Language (TSL) | 2007 | Evaluation results indicates that the model outperforms IBM Model# | No animation or 3D avatar as SL output | TSL glosses |
| [95] | The pre-processing (uses word-tag list) module of the system is incorporated into phrase based and SFST architecture of the system | Spanish to LSE | 2011 | The pre-processing model reduces variability in source language | No animation or 3D avatar as SL output | LSE glosses |
| [112] | Involves sense based and pronunciation based translation and a corpus of Japanese proper names is also constructed | Japanese to Japanese Sign Language | 2014 | Proper names are presented as CG animations | Only human hands and fingers are used to present SL output | 3D model of human hands and fingers |
| [3] | Uses rich module of semantic interpretation, language model and support dictionary of signs | Arabic to ArSL | 2017 | 3D avatar is based on character behaviours and gestures used by ArSL users | Lack of synchronization between facial expressions, lip movements and hand movements | 3D avatar |
| [110] | A word based translation process, uses IBM models for word alignment | English to Indian Sign Language | 2017 | No dependency on rules or grammar of either of the languages | No 3D animation as SL output | ISL glosses |
| [23] | SMT is applied on small corpora giving out satisfactory results | Turkish to TID | 2019 | Produces satisfactory results with small corpora | No 3D animation as SL output | TID glosses |

**Table 7** Projects using Hybrid Machine Translation and their comparative analysis

| Project | Combination | Description | Language | Year | Strengths | Limitations | Sign Synthesis |
|---|---|---|---|---|---|---|---|
| [162] | RBMT + SMT | Phrase structured trees are used along with CFG rules to perform the translation | Chinese to Taiwanese Sign Language (TSL) | 2007 | Evaluation results indicate that the model outperforms IBM Model3 | No animation or 3D avatar as SL output | TSL glosses |
| [115] | EBMT + SMT | Focuses on Irish Sign language grammar and linguistics, also highlights the importance of native signers for manual evaluation | English to Irish Sign Language | 2007 | Sub-sentential chunking of data improves translation accuracy | No animation or 3D avatar as SL output | Annotated output |
| [113] | SMT + EBMT | The decoder of the system compares the input data with the bilingually aligned resources | English to Irish Sign Language | 2008 | Addition of avatar component | Naturality of avatar need improvement | Avatar using POSER animation software |
| [132] | RBMT + EBMT + SMT | Translation module has a hierarchical structure using all the MT systems in each step | Spanish to LSE | 2012 | Field evaluation is adopted to measure the real-time efficiency of the system | Naturalness of avatar is not comparable to human sign language | VGuido 3D avatar |
| [117] | EBMT + SMT | Adds Animation module to the previous project | English to Irish Sign Language | 2008 | Incorporates additional modules over the baseline system | Manual evaluation of signed output is not scheduled | 3D avatar |
| [97] | EBMT + SMT | Only data-oriented strategies are used in this system | Spanish to LSE | 2013 | Sign editor has options like predefined positions and orientations, which reduce sign creation time | – | 3D avatar |
| [99] | EBMT + SMT + RBMT | Technology developed in the previous work has been adapted to a new domain | Spanish to LSE | 2013 | Translation system is developed for two separate domains | Inclusion of RBMT makes translation a time consuming task | 3D avatar |
| [94] | EBMT + SMT | Analysis of different translation strategies and their integration to achieve the best accuracy | Spanish to LSE | 2014 | Extensive field evaluation results in real time effectiveness of system | Lack of normalization of LSE leads to several sign mistakes | 3D avatar |
| [89] | RBMT + SMT | Creating a large parallel corpus that is further used as training data | Greek to GSL | 2018 | Process does not need deep grammar knowledge of GSL | No use of animation technologies for sign synthesis | GSL glosses |
| [2] | RBMT + SMT | Building artificial corpus using grammar dependency rules | English to ASL | 2019 | IBM algorithms are enhanced by integrating Jaro-Winkler distances | No use of animation technologies for sign synthesis | ASL glosses |
| [19] | RBMT + EBMT | Translation rules and a database of signs are used for the translation, and proper names are finger spelled | Arabic text to ArSL | 2019 | Combination of linguistic knowledge and database of signs results in better translation accuracy | No 3D animation. No evaluation | GIF images |
| [79] | RBMT + SMT | After the application of rules, the intermediate result is fed into the statistical component | Turkish to Turkish Sign Language (TID) | 2019 | Language-specific rules increase the overall performance of the system | No 3D avatar or animation for SL output | TID glosses |

**Table 8** Projects using Neural Machine Translation and their comparative analysis

| Project | Description | Language | Year | Strengths | Limitations | Sign synthesis | Dataset |
|---|---|---|---|---|---|---|---|
| [103] | NMT system wholly based on attention was used in this video to video system | English to ASL | 2018 | Combines machine learning and deep learning | Tokenization errors due to low vocabulary size | Virtual avatar | 83,618 pairs of sentences [122] |
| [20] | The project uses feed-forward back propagation Artificial Neural Network | Arabic to ArSL | 2019 | Morphological characteristics are utilized to derive maximum information from each word in the input to NMT | Limited database of sentences and signs | 3D Avatar | 9715 pairs of sentences(interrogative, affirmative and imperative) [19] |
| [147] | Recurrent Neural Network Method is combined with Motion Graphs | German to German Sign Language (DGS) | 2019 | Uses Motion Graph to produce sign video frames | Less training resolution fails to produce sign sequences comparable with avatar animation | HD Sign Video sequences | 8257 sequences of weather broadcasts [24], videotaped repeated production of 100 isolated signs [38], multiple videos of BSL sequences [16] |
| [138] | Progressive Transformer uses a counter decoding technique to predict continuous sign language sequences of varying length | German to DGS | 2020 | Uses several data augmentation techniques to improve SLP production | Focuses mainly on manual features of a sign, and under expressed sign production | 3D sign pose sequences | 8257 sequences of weather forecast with 2887 German words and sign language videos of 1066 different signs [24] |
| [136] | Progressive transformer architecture is employed with a conditional adversarial discriminator | German to DGS | 2020 | Appends regression loss with adversarial loss and introduces production of non-manual sign features | Uses skeletal representation of signs | 3d sign pose sequences | 8257 sequences of weather forecast with 2887 German words and sign language videos of 1066 different signs [24] |
| [157] | Uses a sota human motion transfer method | English to ASL | 2020 | Study shows preference of videos over skeletal visualizations | Bad synthesis of the hands | Sign language videos | 60 h of sign language videos [37] |
| [137] | Transformer architecture with Mixture Density Network (MDN) is employed to generate skeletal poses | German to DGS | 2020 | Novel key point based loss improves the quality of hand image synthesis<br>Controllable video generation enables trainin on large and diverse datasets | | Continuous sign language videos | German vocabulary of 2887 words, Sign language videos of 1066 different signs [24] and high-quality sign interpreter broadcast data |

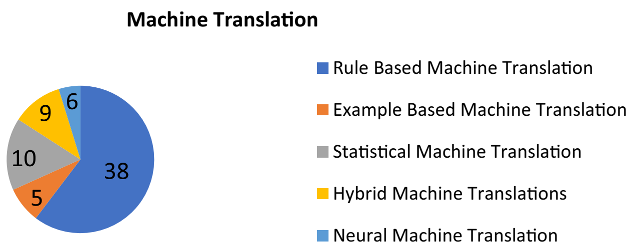



**Fig. 1** Chart depicting machine translation studies

translation via DRT and HPSG, TEAM, SignSynth, ViSiCAST, TESSA [43, 54, 130, 156, 161, 164].

The role of grammar development in sign language synthesis has been an essential task in many earlier works. Safar et al., and Marshall et al., demonstrated semi-automatic translation from English to sign language gesture notation. Marshall et al. further extended it by the development of sign language grammar in HPSG that proposes the uses of a three-dimensional region of signing space for grammatical and discourse purposes in signing [104, 106, 129]. Furthermore, Marshall et al. summarizes these approaches and discuss an extension to the HPSG sign language grammar, which includes SL plurals in combination with the 3D model of a signing space [105]. It was noticed that the synthesis of HamNoSys descriptions within an HPSG framework has been productive. Thus, this was one of the significant advents in virtual signing in SLMT scenario. Later on, Safar et al., and Elliot et al., collaborated for sign language generation using virtual human avatar by using HamNoSys [56] notation which had earlier also proved to be beneficial [43, 50, 130]. All these systems made use of Transfer based or Interlingua based approaches. Huenerfauth's approach, in which he proposed a multi-path architecture, combines all Direct, Transfer, and Interlingua approaches and deals with Classifier Predicates [62]. Huenerfauth discusses the importance of classifier predicates in rule-based systems and how its spatial nature motivates the integration of virtual reality software and provides scope for finding other ASL phenomena using spatial frameworks.

Under the eSign Project's auspices, a significant outset from TESSA was developed under the name of VANESSA [43, 51]. This project was a step up over TESSA as it used a user interface and sign animation avatar for conveying messages between a post office clerk and a deaf client. The overall user response of the system was relatively better than TESSA. The works mentioned above mainly dealt with American Sign Language (ASL) and British Sign Language (BSL) and took the English language as the source language.

Apart from American and British sign language, significant works have been done in other sign languages. The South African machine translation project (SASL-MT) includes translating English input text to South African Sign Language (SASL) [165]. This study uses the STAG parser design to perform the translation and uses H-Anim [173] standard for sign synthesis (Fig. 4). A further study by Zijl et al., under the SASL-MT project, presents that the STAG parser can be extended to include post-processing of the STAG output tree to include meaningful sign space construction and emotive capabilities to improve non-manual sign generation [166]. WordNet [47] is used to tag the STAG trees with form and emotional class information that eases the notation requirement that generates the avatar movement instruction from SASL gloss input.

Zijl et al. further enhance the rule-based approach by integrating the GNApp augmentative and alternative communication (AAC) grid interface [27] with the SASL-MT translation system [165, 167]. This interface helps resolve the fundamental linguistic issues of rule-based approaches

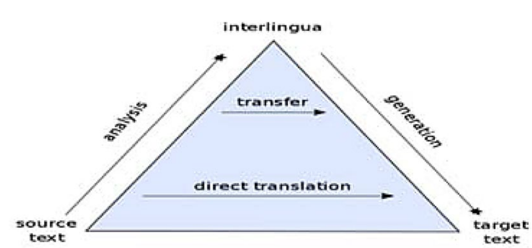
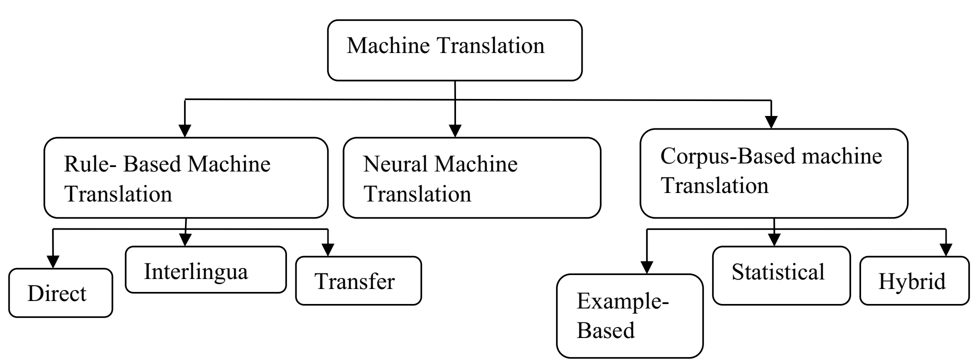



**Fig. 3** Vauquois Pyramid [87]

**Fig. 2** Taxonomy of machine translation studies



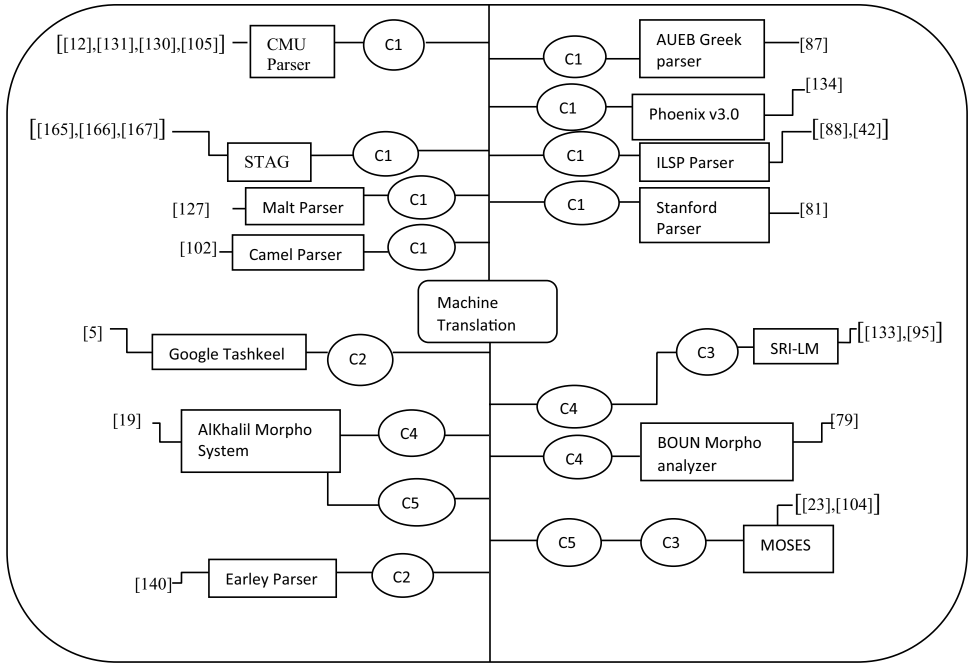



**Fig. 4** Different parsers/analyzers used in machine translation studies

like dependency on perfect POS tagging and word sense disambiguation (WSD) [111].

San-Segundo et al. presented one of the first experiments of translating Spanish speech to a sign language translation system [131]. It followed a limited domain (Identification document, passport) of 458 words and 270 gestures. The sign gestures were represented using VGuido (the eSIGN 3D avatar) by taking an input script in the form of SiGML (Signing Gesture Markup language). The study's evaluation resulted in a 27.2% gesture error rate and 0.62% BLEU (Bilingual Evaluation Understudy), which was due to a high percentage of deletions and other errors. Montero et al. also presented an architecture for translating Spanish into LSE gestures [133]. This work includes four modules: speech recognizer, semantic analysis, gesture sequence generation, and gesture playing. The central focus of the work is on the last two phases. For gesture animations, an animated agent is developed, and the agent positions created by the developer are combined with positions generated automatically by the system. Evaluation for the proposed strategy is done using position distance metric and measurement of gesture complexity. Manual evaluation has not been conducted, which is also marked as a future endeavour in which perception of deaf people will be considered. For Spanish to LSE translation, there is another rule-based approach proposed by Baldassarri et al., which also has different modules for translation of phrases and sentences: morphological analyzer, grammatical analyzer, morphological transformer, grammatical transformer, and sign generator [9]. The signs are generated on the Maxine animation engine, which also includes the mood preferences of a deaf person, which is a step up from the previous work [10].

Regarding teaching SL grammar, Kouremenos et al. presented a prototype Greek text to Greek Sign Language (GSL) conversion system, which is integrated into an educational platform to address the needs of teaching GSL grammar [88]. Kouremenos et al. present the language processing component's detailed implementation focusing on SL grammar knowledge problems in this work. In the same language pair, Fotinea et al. proposed a dynamic combination of linguistic knowledge and avatar performance [46]. GSL is annotated using HamNoSys and VRML controlled by STEP (Scripted Technology for Embodies Persona) language [60] is used for avatar animation. Manual and non-manual features like eye gaze and facial expressions have been considered, but the avatar lacks naturality.

Later, systems using rule-based approaches followed sophisticated parsers and visual avatars for sign synthesis. Mazzei et al. use dependency parser, semantic interpreter, and spatial planner to treat hands position while sign generation [109]. The systems evaluate and compare the results with a statistical translator using Mean and Standard Deviation, giving out approximately the same results. Another rule-based translation approach proposed by Porta et al. for translating Spanish to Spanish Sign Language glosses addresses issues like lexical gap using lexical-semantic relationship, topic-oriented analysis using word order algorithm, and classifier predicates and classifier names [126]. The system is evaluated, giving BLEU as 0.30 and Translation Error Rate (TER) as 42%. The paper's linguistic error analysis indicates that the difference in the system output and reference translations arises from the variations in the linguistic structures and that classifier predicates are the most complex expressions to be generated. However, the paper is incomplete as the translation system gives out glosses as the output, which is considered an intermediate symbolic representation and not a final output.

Almeida et al. use a rule-based pipeline with a deep structural transfer and analysis up to the semantic level to better cope with the language gap [4]. This system has also explored various challenges in the domain of Avatar animation. It used Blender to setup a character with a *rig* move mechanism to generate a fluid animation of Portuguese Sign Language (LGP) utterances [169]. Though the results were positive, the deep linguistic study of sign language was a future concern.

In sign generation or sign synthesis, Braffort et al. developed a web application named KAZOO for French sign language (LSF) generation using a virtual signer [17]. The author describes the platform's architecture in detail, and it discusses the WebGL technology used in it to display 3D animations of readymade animations together with synthetic animations based on SL specific linguistic model [178]. This research uses AZee as the linguistic model that allows more flexibility, precision, and completeness and pays attention to manual and non manual articulators and combines them effectively [49]. Filhol et al. further use KAZOO in another French to LSF translation system based upon two distinct efforts: formalizing LSF production rules and triggering them with text pre-processing [48]. The AZee frameworks use the concept of "production rule" identified by its function. They formally describe the form to produce to express it to be unambiguously generated by SL synthesis with a virtual signer. This approach's second distinctive effort deals with breaking the entire text into as many text processing problems as there are rules available. Therefore each established rule's function creates an information extraction problem. This approach creates numerous amounts of rules to make the system linguistically robust. On the other hand, a large set of rules leads to the problem of rule combination, thus creating confusion of how the rules are to be nested under one another.

RBMT approaches still hold a relevant place in the text to SL translation due to the lack of parallel corpora required in the modern machine translation systems. Lack of grammar and lack of Sign language corpus are the two problems addressed by Kouremenos et al. To further improve upon them, propose a processing method for creating a large and good quality parallel data for SL's [87]. Kouremenos et al.'s system is an advancement over a Greek to GSL translation system of Efthimiou et al. It uses an open-source Python NLTK toolkit than Java technologies and open-source AUEB's Greek POS parser [86] over Greek ILSP parser, which was not open source [42]. The proposed system's gloss system is based on the Berkley gloss system and contains non-manual components information and can be directly used for any 3D animation. The performance of the system is evaluated on a BLEU metric score.

In the Turkish language, Eryiğit et al., a rule-based approach, performs translation for Turkish primary school educational material into TID (Turkish sign language) [44]. It follows a transfer-based approach in which the translation stage consists of translational rules from Turkish to TID. The input to the translational rules stage is analysis of the source language which is produced by Turkish NLP pipeline [45] and the output is fed into the animation layer. The animation signs are collected using a motion capture scheme, which uses RGB-D cameras to capture signs from native signers. Resource and corpus creation of TID is the main focus for which it introduces a machine-readable representation scheme of TID, which is linked to the ELAN annotation tool. The work is part of an ongoing research project which is not yet complete; thus no evaluation is reported.

Luqman et al. develop a gloss notation system to transcribe Arabic Sign Language (ArSL), also creates a semantic RBMT to translate from Arabic to ArSL [102]. The translation process follows three main stages of morphological analysis, syntactic analysis, and ArSL generation. Input to the system is an Arabic sentence, and output is an ArSL sentence represented in a gloss notation displayed as a sign sequence of GIF images. A bilingual parallel corpus targeting the health domain of 600 sentences was build and was translated into ArSL by two expert signers. The system generates 15 rules which cover the mapping at word, phrase, and sentence level. The Out of vocabulary (OOV) problem is also handled using the synonyms of the OOV words. The results reported by evaluation parameters BLEU, WER, TER are 0.35, 0.55, and 0.53, respectively. The cause of errors is mainly because of using a parser whose training has been done on news data and thus is not appropriate for the proposed work, which deals with the health domain data.

Conversion of sentences to gloss notations is converting source language standard sentences into short sentences for the deaf. Nguyen et al. propose a rule-based approach that reduces or shortens spoken or written Vietnamese sentences by reducing propositions, conjunctions, and auxiliary words and replacing synonyms [120]. The system evaluates 200 simple sentences giving out a BLEU score of 97.5%, thus proving the proposed method's effectiveness. However, the system did not perform sign synthesis, which was incorporated by Da et al., which uses a machine-learning decision tree (ID3) to convert Vietnamese sentences to short sentences of Vietnamese Sign Language [29]. In this project, a system is developed that builds a data set and applies ID3 to convert sentences' structure to reduced SL forms. HamNoSys notation is used for transcription and then a SiGML code to express the SL by a virtual signer. Manual evaluation of the results indicates that the understanding of the clip generated is 97.06%. As in all RBMT approaches, the need for a more extensive dataset is the limitation of this project.

One of the latest applications of RBMT is a Pakistan Sign language project (PSL). In this project, Khan et al. use a grammar-based MT model to translate English sentences into equivalent PSL sequences using core NLP techniques [81]. Khan et al.'s project is one of the first projects in PSL using core NLP techniques. This approach analyzes the linguistic structure of PSL and formulates the grammatical structure of PSL sentences. The rules created by analysis are formalized into Context-free Grammar, used as a parsing module for translation and validation for target PSL sentences. Before the generation of rules, a dataset with the deaf community and PSL experts' help was created. The created dataset is then extensively analysed for grammatical differences between English and PSL. Only the sentences present inside the sign database are animated for sign generation, and the rest are fingerspelled. In manual evaluation, valid and invalid sentences were used as parameters. The automatic evaluation was done based upon BLEU, WER, and TER as the performance metric. The respective scores for all three metrics are 0.78, 0.10, and 0.15. The authors also discuss several points that can be incorporated in the future; one is to use the Deep Learning approach to generate more data. The rule-based approaches try to model linguistic knowledge to formalize rules, allowing processing data from the input source to the target languages.

As reviewed above, many distinctive projects have used a rule-based approach because it generates accurate results for small-sized datasets. Table 4 lists all the studies regarding RBMT discussed above and their comparative analysis.

#### 4.2.2 Corpus-based machine translations

Corpus-Based Machine Translation (CBMT) is generated based on the bilingual text. Though the RBMT systems can produce efficient translations, constructing the whole RBMT system is a laborious task as linguistic resources need to be handcrafted, and new rules to be added to the system from time to time, making it a time-consuming task. CBMT systems, on the other hand, are based upon a large amount of bilingual data. CBMT systems, also sometimes called data-driven machine translations, are classified in three categories: (1) Example-based machine translation (EBMT), (2) Statistical machine translation (SMT), (3) Hybrid based machine translation (HBMT).

#### 4.2.3 Example-based machine translation

Makoto Nagao first suggested Example-Based Machine translation in 1984. Training of EBMT is done on bilingual parallel corpora containing sentence pairs of both languages [118]. Sara Morrissey and Andy Way have been pioneers in using the example-based approach in SLMT systems. In one of the first approaches, Morrissey et al. use the Marker Hypothesis [52] to translate English to Dutch Sign Language [114]. The Marker Hypothesis proved to be a promising approach for the segmentation of English input text into chunks that can be aligned accurately with SL annotation. ELAN annotation tool has been used for sign language corpora, including richly annotated data of three different sign languages, including Dutch Sign Language. Improvement was expected in the project's chunk alignment segment to make close matches between the English text and sign annotation.

Almohimeed et al. propose that EBMT is suitable to produce reasonable translation output even with existing small size corpora [5]. It uses a corpus of 203 signed sentences for conversion from Arabic text to Arabic Sign Language in the domain of instructional language typically used in deaf education. The EBMT system works on chunks taken from the input text and aligned with the equivalent signs. Google Tashkeel is used in the pre-processing step to avoid ambiguity. In the final step, the signs are recombined, and the whole sentence is delivered using Window Media Video (WMV). The evaluation results indicate a word error rate (WER) of 46.7% and position-independent word error rate (PER) of 29.4% using the Leave-One-Out Evaluation Technique. The higher error rate is due to the translation of those sentences that are not similar to the examples because EBMT depends on the example's quality. The results are expected to improve with a larger corpus.

Boulares et al. combine EBMT with genetic algorithm and fuzzy logic to translate English into ASL [15]. This approach performs a global proximity search (Needleman-Wunsch Algorithm [119]) to perform global alignment on two sequences and then proceed to Example proximity search (Smith-Waterman Algorithm [160]) for local alignment. This approach results in a set of scores representing proximity between all words in the two sentences. Boulares et al. use fuzzy logic concepts to detect compound emotion from the text, which yielded good results that can be used further in systems that use interpolations between different facial expression modules to produce emotions [15].

A bidirectional EBMT approach for Turkish to Turkish Sign Language (TSL) proposed by Selcuk-Simsek et al. prefers EBMT over other corpus-based approaches as the grammar of TSL is barely known, and also EBMT is suitable for a limited dataset [139]. Learning and translation are two primary components of this system, and both procedures use lexical supervision component (LSC) as its subpart. LSC is constituted of a morphological analyzer, an orthography control tool, and a disambiguation tool. Furthermore, k-fold cross-validation determines training and test sets, and results are obtained in terms of BLEU as 43% and TER as 38%. A significantly lower TER rate was observed by Quach et al. in the process of converting Vietnamese Grammar to Vietnamese Sign Language (VSL) [127]. A corpus of 740 sentences was used, and a TER value of 2.58% was observed. However, the system's processing time was long as EBMT systems translate based on available sentence patterns. EBMT approaches mentioned above are usually applied in projects having a limited dataset. Table 5 lists all the studies discussed regarding EBMT along with their comparative analysis. For larger datasets, this approach is unsuitable as it is challenging to create a large number of examples. Thus for larger datasets, we will discuss the implementation of Statistical Machine translation.

#### 4.2.4 Statistical machine translation

Statistical translations are another form of Corpora Based Machine Translation (CBMT) that works on bilingual text paradigm. The Statistical translation is based on probability distribution which was first approached with Bayes Theorem. Statistical translation, unlike RBMT, does not require manually developed rules, and unlike EBMT, is not suitable for small corpora but is only efficient with large bilingual corpora. Earlier works of Statistical Machine translations include Koehn et al. [84, 85], wherein the former works on word alignment and a framework that enables to evaluate and compare various phrase translation methods. The latter presents a suite of open-source toolkit that, along with an SMT decoder, includes a wide variety of training and tuning tools. Bungeroth et al.'s work is also one of the earlier attempts of SMT to translate German text into German Sign Language (DGS) [22]. In this approach, a statistical scheme modified from IBM models [110] is proposed. Corpus was an issue in this work as only 167 sentence pairs for training and 33 for testing in DGS and German were investigated. The evaluation results of 59.9% word error rate (WER) and 23.6% position-independent word error rate (PER) indicated improvement of the referenced baseline model, but it was considered a small-scale example.

Stein et al. proposed another German-DGS translation system [146].This paper's pre and post-processing steps based upon morpho-syntactic analysis of German are included to enhance the machine translation. For the preprocessing steps, CG parser was used, POS tags were used, splitting of words at breakpoints, and deleting commonly unused words in DGS. Post-processing steps try to curtail basic errors of translation algorithms. Results are measured using WER and PER, which signify a 9% improvement over the baseline system. Corpus building has been a task in many SMT research. Freksa et al. describe the first stage of corpus building and a translation system based on phrase bilingual dictionary for Czech text to Czech Sign Speech synthesis giving a sentence error rate of 50.5% [74]. Another Czech text-to-sign synthesis system proposed by Krnoul et al. converts written text to an artificial human model animation [92]. The translation system implements a simple monotonic phrase-based decoder (SiMPAD) [75] which does not have a reordering module and uses a trigram language model. SLAPE [91] editor is used to create and edit signs, and HamNoSys is used to represent intermediate signs which are further synthesized in H-Anim standard animation model. Dangsaart et al. presented the Intelligent Thai text-Thai sign translation ($IT^3STL$), which follows reordering rules and Thai-sign dictionary to convert a Thai text to Thai Sign language (TSL) [30].It reaches an *F*-score of 97% for 297 sentences for language learning which used knowledge basis and architecture of Thai-Thai Sign Machine Translation (TTSMT) to enhance Thai sign language learning [31].

Data sparseness has been a significant problem in statistical machine translation systems discussed above. Su et al. introduced thematic roles to counter the problem of data scarcity for Chinese to Taiwanese Sign language translation [148]. Thematic roles attempt to capture similarities and differences in the verb meanings, e.g., agent (the 'doer' or instigator of the action) [175]. The authors adopted Synchronous Context-Free Grammar (SCFG) instead of PCFG to convert a Chinese structure to a corresponding Taiwanese SL structure. Translation memory, which comprises thematic role sequences of both languages, contains the learned templates from which the bilingual corpus can be applied for thematic role sequence matching. Another important effort of this approach was the determination of verb agreements to produce expressive sign sequences. BLEU scores achieved by the system for long sentences (with $n$ values of 3 and 4 in $n$-gram precisions) are 0.65 and 0.671, respectively. However, due to little research on Taiwanese SL linguistics, non-manual features are hardly constructed in this Structural statistical approach.

Lopez-Ludena et al. proposed to improve in Statistical methods of machine translation in their study. The approach in this paper analyzed two different statistical strategies: Phrase-Based system (Moses [84]) and Statistical Finite State Transducer (SFST) [100]. Two new methods are used to improve translation from Spanish to LSE. The first consists of implementing a categorization module (which replaces Spanish words with associated tags) in the preprocessing step before the translation. The second is the use of Factored Translation Modules (FTMs) for improving translation performance. The first method has been incorporated into the Phrase-based system and an SFST, but FTM is considered only in the Phrase-based system. These methods allow incorporating syntactic-semantic information during the translation process, thus reducing the source language variability and number of words composing the input sentence. The valuation module reveals that the categorization module's use increases BLEU from 69.1 to 78.8% for phrase-based systems and from 69.8 to 75.6% for SFST outputs. The inclusion of FTMs also increased the BLEU score from 69.1 to 73.9%. Lopez-Ludena et al. did a step up from this work, in which the pre-processing module replaces Spanish words with associated tags and removes the words having 'non-relevant' tags [96]. The pre-processing module was incorporated in a phrase based and SFST statistical translation system. A parallel corpus of 4080 Spanish sentences and their LSE translation has been used, and the valuation results of BLEU rose from 73.8 to 81.0% in phrase-based system and 70.6% to 78.4% for SFST. The paper has also presented a human evaluation (two experts); pre-processing module obtains an increase of 0.64 to 0.83 in phrase-based and an increase from 0.65 to 0.82 in SFST.

Data collection is also a significant component of SMT systems. Stein et al. analyse existing data collections and emphasises their quality and usability for statistical machine translation [145]. This work analyses different existing corpora like RWTH-Phoenix corpus [120], which is a richly annotated corpus, Corpus-NGT [121], SIGNSPEAK [122]. The second part of the project deals with the preparation of Sign language corpora for DGS and translation from German to DGS. Sentence end markers are introduced in the pre-processing phase of the translation process. The last phase of the paper discusses the optimization of the scarce resource translation procedure. The results indicate whether the project's domain was suitable for the machine translation applied, but still, it had less impact and usefulness for the deaf user.

Miyazaki et al. proposed a Japanese proper name translation system that involved sense-based and pronunciation-based translation [112]. The sense-based translation is learned from phrase pairs in a Japanese-JSL (Japanese Sign Language) corpus, and pronunciation based is learned from a Japanese proper name corpus. The corresponding CG animation, a high-quality 3D model of human hands and fingers, is created when a proper name is entered. The CG animation is rendered from scripts written in TVML (TM program Making Language).

Al-Barahamtoshy et al. use a rich module of semantic interpretation, language model, and support dictionary of signs to understand the type, tense, number, gender, and the semantic features for subject and object, which will be scripted by a 3D avatar [3]. This approach uses SMT for alignment, which is considered a function of transformation between source and target language. The future implications of this work include acquiring facial expressions and lip movement synchronized with hand orientation.

In the domain of ISL (Indian Sign Language), Mishra et al. point the limitations of the Dasgupta et al. approach, indicating that the latter was not a generic model to be followed for machine translation [32, 110]. Mishra et al. translated English to ISL glosses using a word-based translation model, and the methodologies are implemented on MOSES [84]. The corpus consists of 326 English sentences and 537 ISL glosses. Integration with phrase-based translation for better results is the prospect of the paper [110].

In the Turkish language, Buz et al. presented a novel approach to convert primary school education book material to TID using SMT [23]. Earlier, Eryigit et al. attempted the same language pair translation, following the RBMT approach [44]. In this, five different approaches have been tested, every approach applying a different kind of pre-processing to the source data. In the first approach, no operation was performed on the data and gave the BLEU score of 61.69% and WER to be 42%. In the second approach, Stemming was applied, which improved BLEU-1 to 77.66%; for approaching third, positive and negative tags were given to each verb to contain their meanings while stemming. In approaches 4 and 5, pronouns were added after examining verbs, leading to a drop in BLEU-2 and BLEU-3. Though the system indicated an excellent SMT approach even for smaller corpora, the system lacked any manual evaluation and visual synthesis of signs. Table 6 lists all the SMT systems considered in this review. SMT and EBMT approaches do not require high-end linguistic knowledge, but they suffer from the limitation of parallel data's unavailability. Thus several researchers have tried combing corpus-based and rule-based strategies to achieve better results and response.

#### 4.2.5 Hybrid Machine Translation System

Multiple machine translation systems within a single machine translation system are Hybrid Machine Translation (HMT) systems. The need for a hybrid machine translation system arises from the failure of single machine translation systems to achieve an adequate accuracy level. For example, Hogan et al. at Carnegie Mellon University combined example-based, transfer-based, knowledge-based (a rule-based system displaying extensive semantic and pragmatic knowledge of domain), and statistical translation sub-systems into one machine translation system [59].

Wu et al. combine rule-based and statistical approaches to achieve translations from Chinese to Taiwanese Sign Language [162]. The authors use a parallel corpus of 2036 sentences of Chinese and a parallel annotated sequence of corresponding Taiwanese Sign Language words from which CFG rules are created, and transfer probabilities are derived. Context-Free Grammars (CFGs) are formal grammar designed for transfer-based statistical translations. Another corpus, a Chinese Treebank containing 36,925 manually annotated sentences, is also used. Both corpora's are used to derive a probabilistic CFG. The system produced a BLEU score of 0.86 and was also manually evaluated with good, fair, and poor three traditional opinions. Though the results generated were satisfactory, the system suffered from small-sized corpora problems consequentially unsuitable for statistical translations. The use of rule-based approaches inhibits the system of extensibility to new language pairs. Also, the system did not produce any signing avatar.

Researchers have tried using HMT systems to improve practical applications of SLMT. Morrissey et al. uses the MaTrEx Machine translation system [121] and combines SMT and EBMT methodologies to make a modular design that is above all adaptable to convert English to Irish Sign Language (ISL) [115]. The results indicate that the system does a good work for conversion but loses its practical use as it does not have a signing avatar. Morrissey improves this system in the following study [113]. In this approach, the MaTrEx decoder is fed with three bilingual data resources: aligned sentences, sub-sentential chunks, and words. For sign synthesis, POSER [170] animation software version 6.0 is used to create an animated avatar. Though the system used the concept of animation, the avatar lacked naturalness in its movement and expressions.

Morrissey et al. addressed corpus-driven MT predominantly for Irish Sign Language and DGS [117]. They extended the MaTrEx system by adding two additional modules of recognition and SL animation. Though SL recognition is out of our survey's scope, SL transcription and evaluation are other challenges handled in this work. The approach conducts two experiments using two different corpora. Air Traffic Information System (ATIS) corpus [21] is used in the first experiment, and POSER animation software tool is used for producing 3D human figures. In the second experiment, a medical corpus is used, and HamNoSys is used for annotation. Both manual and automatic evaluations have been conducted in the experiments.

San-Segundo et al. develop a comprehensive HMT approach wherein they initially implement and evaluate rule-based, example-based, and statistical translators. The final version combines all the alternatives into a hierarchical structure [132]. A corpus of 2214 Spanish sentences was used, and two LSE experts converted to LSE. For every module in this paper, different strategies are used; the first consists of EBMT, wherein the translation process is carried out based on the similarity between a sentence to be translated and examples of the parallel corpus. The second follows a set of translation rules for rule-based translation and for the last strategy of statistical translation, parallel corpora are used for training language and translation model. Finally, the alternatives mentioned above are combined into a hierarchical structure. The system's evaluation is conducted based on objective parameters like Sign Error Rate (SER), position-independent word error rate (PER), and BLEU. The results indicate that RBMT systems obtain better results than EBMT and SMT systems because the rules introduce translation knowledge not seen in the parallel corpus. The combination of techniques improved the results manifold relative to the single technique results. The paper also depicts the sign animation using VGuido [168] in eSignEditor [57].

Lopez-Ludena et al. presented a more refined version of the approach in which instead of following all data-oriented strategies, the required modules are generated automatically from a parallel corpus [97]. The statistical translation includes a pre-processing module [95] that increases the performance. In this step-up project, the combination of EBMT and SMT is used, and the rule-based strategy module is not included. The sign editor uses inverse kinematics (IK) which helps in reducing the sign specification time. The whole system presents an SER 10% lower than the San-Segundo et al. approach [132]. Lopez-Ludena et al. again used the combination of EBMT and SMT in two new domains of hotel reception and bus information [99] [94]. In the hotel reception domain paper, Lopez-Ludena et al. use a declarative abstraction module for all internal components and H-Anim for sign generation [99]. The paper's objective evaluation was done using WER, SER, and average Translation time with respective scores as 6.7%, 10.7%, and 3.1 s. Subjective measurement is done in the form of questionnaires. This approach's main disadvantage is that the methodology is sequential, and the technology adaptation depends upon the parallel corpus generation. Lopez-Ludena et al. follow the same translation strategy for the bus information domain but focuses majorly on the sign synthesis part [94]. Non-manual signs are also considered during sign language generation. Another advantage of the representation module is an adaptation to different kinds of devices (computers, mobile phones, etc.). The subjective evaluation of the paper involves two main aspects of intelligibility and naturalness. Automatic evaluation metrics indicate SER less than 10%, BLEU greater than 90%, and translation time as 8.5 s. Kouremenos et al. proposed a novel prototype system to create a parallel corpus using RBMT and then use the same corpora as training data [89].

A professional translator, with the help of RBMT, produces a high-quality parallel Greek text to GSL glossed corpus of 1,015 sentences and 20,287 tokens. RBMT uses different tools and technologies: AUEB's POS parser, NLTK 3.0 suite, and Java and Perl scripts. Finally, a parallel corpus trains the MOSES SMT system. When evaluated using BLUE n-gram, the results indicate that the larger the n-gram superior is translation accuracy. On similar lines but without using a professional translator, Achraf et al. created an artificial corpus using grammatical dependency rules [2]. The corpus was of English-ASL language pair and was fed into an SMT system. Both systems mentioned above produced gloss output and expressed the need to incorporate 3D animation in the future.

Brour et al. formulated a hybrid approach ATLASLang MTS by combining EBMT and RBMT to facilitate translation from Arabic to ArSL [19]. In this hybrid approach, if the sentence to be translated exists in the database, EBMT is applied, and if it does not, then the rule-based Interlingua approach is applied. For the pre-processing of Arabic text, the Alkhalil Morpho System [1] is used as an analyzer, and SAFAR Platform [143] is used to transform the analyzed text in the output from html to xml format. When the sentence exists in the database, it is displayed directly without any analysis, but if it does not, then after morpho-syntactic analysis, eleven rules of syntactic reordering are applied. A database of gif images is used to display the final output, and if a sentence contains a proper noun, it is animated using a 3D hand. The final version of translation must use a 3D avatar instead of a gif.

Kayahan et al. observe that language-specific rules tend to increase the overall system's performance and system's work efficiently when combined with other machine translation systems [79]. This system converts Turkish spoken language to Turkish Sign Language by combining rule-based and statistical machine translations. The whole system is divided into three components; the first one is a rule-based translation component which is python-based and comprises 13 translational rules. The second component is a pre-processor that reduces data thinness for the next segment, a statistical translation component. The SMT component of the system uses MOSES for generating a language model and decodes the input sentence. The approach uses the BLEU metric for evaluation and reports the scores as BLEU-4 12.64% BLEU-3 19.28% BLEU-2 31.48% BLEU-1 53.17%. The results reported are satisfactory; however, the system needs a virtual avatar tool for completeness.

Table 7 lists all hybrid translation strategies used in this study. A large amount of work in the text to sign language translation has been done using the conventional machine translation approaches. Since the advent of neural networks, researchers worldwide have focused on including neural network technologies in this field.

#### 4.2.6 Neural machine translations

The Neural Machine Translation (NMT) model uses an artificial neural network to predict the likelihood of a sequence of words. They require lesser memory than SMT systems, and they do not use separate language models, translation models, and reordering models but are just one integrated model. It uses deep learning and representation learning to perform translations. The basic idea behind NMT is to code a sequence of variable length words into a fixed-length vector that summarizes the complete sentence [20]. NMT methods for translation from text to sign language are still unexplored, and an open problem, a few but significant NMT works are discussed in this part. As deep learning and Artificial Neural Network technologies took advancement machine translation methods, Manzano et al. used NMT to translate English to ASL [103]. The approach worked on the ASLG-PC dataset [122], which consisted of 83,618 pairs of sentences, was used for translation and ASL glosses were the output of the process. The vocabulary size of the project was small, which led to tokenization errors.

A major translation from Arabic to Arabic Sign Language ATLASLang had earlier used classical machine translation methods that suffered from the limitation of linguistic knowledge necessary to develop the rules [19]. A dataset of around 9715 input–output pairs of Arabic-ArSL examples of sentences trains the ALTLASLang MTS1 system. The process starts with morpho-syntactic analysis wherein each word is given morphological characteristics, and then the sentence is encoded. On the target side, a target vector is generated using a feed-forward neural network with backpropagation. In the end, the vector produced is decoded using a 3D avatar. For sign generation in the form of 3D avatar XML encoding of HamNoSys called SiGML was developed. By converting HamNoSys symbols to SiGML form, all signs have been displayed using JASigning API. ATLASLang NMT outperformed ATLASLang MTS, with the forming scoring a BLEU score of 0.79.

Another NMT system, Text2Sign, uses a Generative Adversarial Network and Motion Generation to produce sign videos from spoken language sentences [147]. The project is divided into phases, and in the first phase, Recurrent Neural Network (RNN) [26] method of NMT using Luong attention [101] is combined with Motion Graphs [90] to generate sign pose sequences. The resulting pose is used to condition a generative model to produce video sequences. Multiple datasets have been used in this approach:PHOENIX 14 T German weather broadcasts dataset [24], SMILE dataset [38] to train multi-signer generation network, and HD dissemination material acquired by the Learning to Recognize Dynamic Visual Content from Broadcast Footage (Dynavis) project [16] to train HD sign generation network.

The use of multiple datasets demonstrates the robustness of the system. This approach counters the use of avatars and motion capture as for avatars, deep expert knowledge of animation is required, and the Motion Capture method is an un-scalable and costly process. Despite this, the system cannot compete with the existing avatar approaches due to low translation training data resolution.

Contrary to the Stoll et al. approach, which used gloss as a priori to generate sign sequences, the approach proposed by Saunders et al., focuses on automatic sign language production and learning the mapping between text and sign pose sequences directly [138]. This approach used transformer architecture and produced 3D sign pose sequences as the final output. The approach increased the SLP performance when evaluated on PHOENIX-14 T dataset. However, the approach focused mainly on hand and body articulators, thus ignoring a sign's non-manual features.

As non-manual features provide contextual and grammatical information of a sign, Saunders et al. included NMF's in their adversarial multi-channel SLP approach [136]. This model fully encapsulated all sign articulators, thus generating realistic sign productions.

The approaches mentioned above mainly produced skeleton pose sequences that resulted in under articulation, and also no studies have been so far on whether they are helpful to deaf people. Ventura et al. go one step further from skeletal visualizations and generate realistic videos using the state-of-the-art human motion transfer method Everybody Dance Now (EDN) [25, 157]. For signing videos and keypoint, they use a subset of the How2Sign [37] dataset, a large dataset of ASL sign videos. The study results indicate that generated videos are preferred over skeletal visualizations, but the model fails to generate high-quality hand images. Henceforth, the SIGNGAN approach used a Mixture Density Network [14] for more expressive sign production [137]. This approach produced photo-realistic continuous sign language videos directly from the spoken language. The model training was done on separate datasets, PHOENIX-14 T, and a corpus collected from sign language interpreter's broadcast. This system outperformed all baseline systems when evaluated on quantitative metrics and human perceptions.

All of the above discussed NMT approaches are state-of-the-art systems. They have been trained and evaluated on challenging datasets like PHOENIX-14 T. Studies reviewed under NMT are listed in Table 8.

### 4.3 Sign synthesis

An avatar accompanies a complete SLMT system to perform signs. However, specific state of the art systems have not synthesized translation into avatar but have presented the sign words using some notation. Like any other notation system, gloss is also considered a notation system to represent sign language. A gloss is an approximation of a word of another language written in upper case stem form. Stein et al.'s statistical machine translation study uses gloss notation to represent DGS [146]. A rule-based system proposed by Porta et al. uses a transfer approach to convert Spanish LSE glosses [126]. Several other annotation tools help in representing sign languages. Morrissey et al. used the ELAN annotation tool [171], which provides a graphical interface and helps illustrate the corpora in a video format [114]. Almohimeed et al. also use the ELAN annotation to transcribe ArSL in Arabic to ArSL translation [5]. Likewise, many latest approaches have confined their work only till notation representation, though some like IT$^3$STL and ATLASLang MTS approaches have used pictures and GIF images as the presentation mode [19, 30].

Although many SLMT works focus on translation, several comprehensive research pieces have made sign generation a prominent part of the study. In this part of the paper, we will discuss some of the earliest approaches of SLMT, which incorporated sign synthesis as a significant part.

An interlingua approach ZARDOZ describes a multilingual translation system using a blackboard control structure [154]. This system offers a complete generation phase, including a detailed avatar animation phase. Sign tokens are compiled into Doll Control Language (DCL) program by the DCL animator [156]. This program controls an on-screen animated doll to articulate the correct gesture sequence. The TEAM project also uses gloss notation to present the SL output. In the last phase of the project, the sign synthesizer uses an avatar model to show each sign [164]. The TEAM project formed the basis for the translation work carried out in the SASL-MT project [165, 167]. The system differs from the TEAM project as it did not integrate the translation phase with the animation phase.

The ViSiCAST and eSIGN projects produce SL output in HamNoSys notation, DRS, and HPSG [12] and focus on animation generation [107]. Though these projects cater to a wide area of SL animation, it lacks the working of non-manual features.

An ASL workbench focuses majorly on the representation of ASL rather than translation [144]. It adopts the Movement-Hold model of sign notation which divides the signs into sign sequences according to their phonological features. Each phoneme comprises a set of features specifying its articulation. The author noted that a translation system could produce output in many forms like glossed output, linguistic representations, and animated signing, but the project did not have any. However, the study tried presenting a phonetic structure for non-manual features.

The representations mentioned above help in understanding the SLMT. Still, they are not applicable in the real world because deaf people need a system to see the particular sign words' performance by real-life virtual humans as they are much suitable for sign representation. The following part of the paper will discuss the current status of 3D sign synthesis in SLMT.

#### 4.3.1 3D avatar

A 3D avatar as a sign representation mode started as early as the late nineties. ZARDOZ uses a doll animator to perform different articulation of signs. Advancing on the same scheme, TESSA came into play, which used the motion capture method to directly capture human signer movements and coupled this with virtual humans [12]. To produce smooth movements caught using the motion capture technique, TESSA used the Simon-the-Signer 3D model for animation. TESSA formed a stepping stone for another system named VANESSA. VANESSA was one of the several applications based on synthetic signing avatar technology developed under the eSIGN project. In this system, each lexicon entry is expressed in terms of HamNoSys notation which defines manual and non-manual features of a sign. Further, the HamNoSys scripted lexicons are automatically converted to SiGML notation to drive the avatar.

Many projects in the early 2000s focussed majorly on sign synthesis, one of them being the sign synthesis project by Grieve-Smith et al. This project took a sign language text as input in ASCII-Stokoe notation converted it into a linguistic representation further to 3D animation sequence in Virtual Reality Modelling Language (VRML or Web3D) automatically rendered to a Web3D browser [54]. The prototype was one of the initial attempts to create a 3D avatar, thus faced several inverse kinematics problems. The paper also discusses adopting HamNoSys /SiGML or some other linguistic representation for the future.

Kennaway et al.'s synthetic animation system automatically synthesizes deaf signing animations from HamNoSys transcription [80]. A simple control model of hand movement is combined with Inverse Kinematics calculations for placements of arms. The synthesized animation is further combined with motion capture data for the spine and neck to add natural ambient motion. This approach produced results that compared favorably with the existing alternatives of that time.

Though the systems mentioned above focussed on synthesis technology to produce a better performing sign language avatar, there was also a need to exploit natural language processing mechanisms to build structural rules to create a sign-coded lexicon. Efthimiou et al.'s Greek to GSL approach used parsed output of GSL structure patterns enriched with sign-specific information to activate a virtual avatar in Web3D [41]. This approach also introduced an idea and importance of classifiers in sign language conversion systems. Karpouzis et al. present another Greek to GSL system which utilizes language resources suitable for young students to implement a virtual signer software component using VRML plug-in in Web3D browser [77]. Web Sign technology which is based on Web technology has also been majorly used by Mohamed Jemni et al. in his text to sign language approaches [70–72]. All the approaches describe the web based module and its integration with the VRML player. This tool has been used to create sign dictionaries and making information more accessible to deaf users.

Meanwhile, Huenerfauth et al. presented his idea of classifier predicates wherein a 3D visualization of the arrangements of objects in the sentence of English input is done using AnimNL and multimodal NLG technology (a technology for illustrating gestures that are not easily encoded as text strings) [63, 65, 66]. Huenerfauth et al.'s approaches also allowed a feedback evaluation wherein the 3D prototype was compared against earlier used motion capture approaches. Another significant work proposed by Huenerfauth et al. added linguistically motivated pauses and variations in sign duration of avatars created by Sign Smith Studio [177] to improve the signers' performance [61]. The results indicate that signs were more comprehensible and understandable by the deaf users after the proposed changes. Furthermore, Huenerfauth et al. also tested the effect of spatial reference and verb inflection on SL animations [67]. In the experiment's context, 10 paragraph length stories in ASL were designed, and computer animations were scripted using Sign Smith Studio. Native deaf evaluators used Likert scale to answer questions regarding comprehensibility, understandability and naturalness of animations. The study and the results demonstrated that the inclusion of properly inflected ASL verbs produced more realistic animations.

Sign synthesis has been a significant focus in many Spanish to LSE projects. San-Segundo et al.'s approach incorporate avatar use after the system's transfer phase [133]. An animated agent is developed, which is a simple representation of a human person but is detailed to produce accurate gestures required for sign language. This animated agent is a 2D agent named AGR (agent for gesture representation). This avatar had poor level accuracy, as can be attributed to a 2D presentation. Baldassarri et al. adopted HamNoSys and SiGML notation as intermediary representations and then translated them into a signing avatar [9]. This work included a deaf person's mood and used the Maxine Animation engine [10] to present 3D scenes in real-time using a virtual avatar. The use of avatars kept gaining popularity, and people worldwide tried finding new ways to incorporate virtual signer animation in SLMT systems.

KAZOO used two animation frameworks, Octopus [18] and GeneALS [34], to create a specialized animator for animation generation [17]. Gene ALS framework is used to compute postures, and the co-articulation and combination capabilities of Octopus produce final animation.

Another famous 3D sign synthesis standard used in various studies is the H-Anim standard [173]. In this project, each phalanx of the avatar can be positioned and rotated using realistic human animations. Lopez-Ludena et al. used a sign editor module and H-Anim to reduce the time consumed in the generation process of signs [98]. Lopez-Ludena et al. again used the same standard in a subsequent study where manual and non-manual features are composed to produce final animation using Non-Linear Animation (NLA) techniques [94].

Kipp et al. focus majorly on the creation and evaluation of SL avatars [82]. This work has also focused on evaluation and has introduced delta testing as an evaluation method. This technique is a unique way of comparing avatars with human signers. After discussing earlier projects like ViSiCAST, eSIGN, and PAULA (Practising ASL using linguistic animation [33]), the paper proposes the use of EMBR [58] character animation engine as it offers a high degree of control over the animation and is publically available. Evaluation tests for comprehensibility resulted in a score value of 58.6%, which is somewhat close to the state-of-art method ViSiCAST having a 62% score. The paper stresses steering research towards SL synthesis, focussing majorly on non-manuals and prosody to achieve next-level naturalness in avatars.

Another breakthrough project in 3D sign synthesis was the DICTA-SIGN project, which was the first multilingual system and worked on four European languages (Greek, British, German, and French) [40]. The project aims to amalgamate recognition, animation, and machine translation techniques. The signs are synthesised using HamNoSys/ SiGML notation. DICTA-SIGN includes prosodic information along with phonetic and grammatical information. The authors aimed to make DICTA-SIGN a multidisciplinary approach in the future.

As and when researchers became comfortable using 3D animation as a sign synthesis approach, they started moving into the arena of making 3D animation more and more natural. Non-Manual features are capable of carrying crucial linguistic information. Ebling et al. present a work that bridges the gap between the sign translation system's output and the sign animation system's input by incorporating non-manual features at the end of translation [39]. Sequence classification is used to generate non-manual information in sign languages. The glosses generated after translation serve as input to the sequence classification system. Sequential condition random fields (CRFs) [150] are the state-of-art approach to perform sequence classification. The key feature of this paper, cascading of sequential predictions, can be used in other sign languages.

In recent years, the use of free software available for creating 3D avatars started getting popular. There are software's available used for 3D creation like Blender, Unity, and Maya [169, 174, 176]. Blender being open software has been used in many pieces of research. Almeida et al. used Blender software to convert Portuguese to Portuguese Sign Language (LGP) [4]. In the same language pair, a prototype was developed called OpenLibras, which offered a multilingual platform for text-to-sign translation [158]. Open Libras was an extension of VLibras, which followed the same suite of animation [7]. In this system, a 3D avatar model was created composed of 82 bones for facial expressions, hand shapes, body, and arm movements. The model developed has the capability of representing non-manual features.

The above overview of several papers dealing with 3D sign synthesis highlights the importance and advantages of the 3D avatar model over other sign synthesis representations like gloss, videos, pictures, and GIFs because 3D avatars add more naturalness and flexibility in the presentation of SL.

### 4.4 Performance metric in SLMT

An SLMT performance metric is a standard measure of a degree to evaluate translation accuracy from a spoken language to sign language. Evaluation of SLMT is broadly divided into manual and automatic evaluation. Both evaluation methods have their performance metrics that help better understand the accuracy of the system. For manual evaluation feedback from deaf and SL experts, it is one of the best ways to check the understanding of the sign language generation output. Automatic evaluation performance metrics depend upon the type of Machine Translation. In the earlier approaches like ZARDOZ, TESSA, evaluation was not a part of the project [12, 53, 155, 156].

The performance metrics followed by SLMT are more or less the same as that followed by text-to-text MT or transliteration. Word Error Rate (WER) and Sentence Error Rate (SER) are the primary and most widely used performance metrics in transliteration [121] [116]. SER computes the percentage of incorrect complete sentences match. WER calculates distance (Levenshtein Distance algorithm) between reference and candidate translation. The formula of WER is as follows, summing up all kinds of errors (substitution, insertion, deletion).

$$\mathrm{WER} = \frac{S + D + I}{N}$$

where $S$ is number of substitutions, $D$ is number of deletions, $I$ is number of insertions, $N$ is the number of words in reference.

As WER follows the order of the words, another performance metric, PER (Position Independent Word Error Rate), completely neglects the word order [151]. It measures the difference in the count of words occurring in candidate and reference sentences. The resulting number is divided by the number of words in reference.

Translation Error Rate (TER) is another commonly used metric for MT evaluation. It attempts to measure the minimum amount of editing that a human would have to perform to convert the system output into reference translation [142]. The lower TER rate indicates the higher accuracy of the MT system [108]. The TER score is determined based upon the minimum number of corrections Nb(op) on the average size of the reference AvregNref.

$$\mathrm{TER} = \frac{\mathrm{Nb(op)}}{\mathrm{AvregNref}}$$

BLEU [125] is another performance metric that has gained importance in the MT field. BLEU was the first automatic

MT evaluation metric to show a high correlation with human judgment. BLEU score is a precision-based metric that compares a system's translation output against reference translations set by summing over 4-g, trigram, bigram, and unigram matches found divided by the sum of those found in the reference translation set. The score of output translation is produced between 1 and 0. BLEU is a precision measurement; a higher value indicates better translation.

An improvement over BLEU was developed and was called NIST [36]. NIST is also a precision measure, but it avoids BLEU's bias towards short n-gram candidates. This bias is termed as brevity penalty and was an unwanted effect of BLEU [108].

Though the SLMT can use the same metrics, other parameters help measure the quality of the SLMT system's output like San-Segundo et al. use the feedback of 10 users to recognize several gestures played by the system and measured the gesture quality based on it [133]. Similarly, Almeida et al. evaluate an interface created for 3D viewing for usefulness, usability, translation quality, and adequacy of the 3D avatar [4]. Thus, evaluation is divided into two categories: 1) Automatic Evaluation and 2) Manual Evaluation. The next part of the paper will discuss both the evaluation criteria and the importance of each.

#### 4.4.1 Performance metric in RBMT

Cox et al. carried out a detailed evaluation of a translation system developed for communication between a deaf person and clerk in the post office to determine the usefulness of the Wray et al. prototype system for people who have BSL as their first language [28, 161]. The system's evaluation measures the quality of signs, difficulty performing a sign with TESSA, and perception of the deaf users and post office clerks. On a 3 point scale from 1 (Low) to 3 (High), the average ratings of acceptability by deaf persons was 1.9 with TESSA and 2.6 without TESSA. The clerks also rated the transactions completed with TESSA slightly lower than without using it. In conclusion, the average rating was 2.5 with TESSA and 2.6 without TESSA. For subjective evaluation, several questions were asked from deaf participants about ease of communication in the post office. The responses were measured on a 5 point scale from 1 (Very difficult) to 5 (Very easy). Herein the clerks responded that communication was "slightly easier" or "much easier" with TESSA than without it.

VANESSA, a significant project under eSiGN project, was amongst several applications based on synthetic signing avatars [51]. The evaluation conducted in this project was a threefold user evaluation. The first one was on-site with the live VANESSA system with two BSL users. The second was laboratory-based evaluation with 7 BSL users. The last was a simple evaluation of the intelligibility of the sequences performed by the signing avatar by 15 BSL users. Though no automatic evaluation was done, the results of the manual evaluation were well documented along with scale values marked by (1) Liked, (2) Quite liked, (3) Neutral, (4) Not much liked, (5) Disliked and relevant participant comments. Fotinea et al. also subjected its system to evaluation regarding the system's usability and its appeal to the users concerning navigation, educational targets, and virtual signers' acceptability [46]. Evaluators' comments were categorized according to the virtual signer's naturalness, performance accuracy, and appearance.

A study is considered to be a complete one if both automatic and manual evaluations are conducted. Dangsaart et al.'s evaluation system measured translation accuracy (Translation of sentence and sign representation) for automatic evaluation and user satisfaction for manual evaluation. Translation accuracy was measured in terms of intelligibility and fidelity by testers to determine if the system generates correct and reliable translations [30]. The performance metric used in this case were accuracy, precision, recall, and $F$-score. The equations of the mentioned metrics are as follows:

$$\text{Accuracy}\ (Y|X) = \frac{|Y|}{|X|}, \quad (Y, X > 0)$$
$$\text{Precision}\ (p) = \frac{|X \cap Y|}{|Y|}, \quad (Y, X > 0)$$
$$\text{Recall}\ (r) = \frac{|X \cap Y|}{|Y|}, \quad (Y, X > 0)$$
$$F_1(r, p) = \frac{2rp}{r + p}$$

User satisfaction was evaluated in terms of thought and preference of the deaf via questionnaire. The scale taken for the measurement was rates on 5 points (5-Excellent, 4-good, 3-fair, 2-poor, 1-very poor). The evaluation results of the system show that the system performance satisfies the user's need.

Translation speed is another metric that measures the quality of the system. Baldassarri et al. measure the system's performance based upon well-translated words, words in the correct order, and the translation speed [11]. Thus the time required to translate from a spoken language text to the animated sign indicates the translator's relevance to work in the real-time system. Translation speed and TER are also performance metrics used by Da et al. to measure Vietnamese television news translation performance into 3D SL animation [29].

Porta et al. has evaluated the Spanish to LSE translation system using BLEU and TER performance metrics and compared the evaluation with other approaches [126]. It conducts six experiments to conduct a comparative analysis with different methods based on TER and BLEU performance metrics.

The system does not produce any animation; thus, no evaluation for synthesis is done; the study also lacks manual evaluation, which is considered necessary for a complete SLMT. In the ATLAS project, a similar comparison evaluation has been presented but using different metrics. In this project, the mean and standard deviation of the correctly recognizing correct and incorrect paraphrases by proposed and the referent systems have been compared [109].

The manual evaluation applied differs from system to system; on the contrary, automatic evaluation metrics are common to several systems. Luqman et al. present BLEU, WER, and TER as performance metrics for automatic evaluation. For manual evaluation, user feedback from one deaf person and one bilingual translator with good, fair, and poor metrics to measure translation quality [102]. In an Arabic translation approach, Al-Barahamtoshy et al. only conduct the manual evaluation. It evaluates the quality of signs by classifying them into valid, partially valid, or not valid categories [3].

OpenSigns, a multilingual translation platform from multiple spoken languages to Portuguese Sign Language (LIBRAS), was compared with the original VLIBRAS using BLEU and WER performance metrics [158]. The evaluation was conducted in two phases. In the first phase, text-to-text translation was tested using BLEU and WER, and in the second phase, using the same metrics, text-to-gloss translation was tested. OpenSigns was one of the first systems using BLEU and WER metrics to test a multilingual platform. A manual evaluation of the system was conducted using a three-part questionnaire with every part consisting of different questions. The test was performed with 15 deaf Brazilian students and three LIBRAS interpreters from three public institutions. The user had to select between 5 responses. Though animation quality was not tested, this study tried to perform a complete and proper evaluation of the system in all this study.

Evaluation is a challenge when there is a lack of substantial corpus for assessment. In a study based on English text to Pakistan Sign Language (PSL), the authors did a time-consuming task of compiling a sentence level evaluation corpus of 500 sentences, which covers all categories of PSL sentences corresponding to various linguistic types of English [81]. Corpus is evaluated using BLEU, WER, and TER performance metrics. Manual evaluation is also conducted with two deaf subjects and two bilingual experts of PSL and English. Another approach that reduces evaluation time is followed wherein only the translation rules are evaluated and not the manual comparison of each sentence with its transcription. In this work, 820 transfer rules are extracted, and precision is calculated as

$$T\,(\text{precision}) = \frac{\text{count}\,(\text{valid sentence})}{\text{count}\,(\text{sentence})} \times 100$$

We can infer from the above submitted reviews that though manual evaluation is time-consuming, it still holds an important place to test the system's validity, usability, and overall acceptance. Automatic evaluation, a less time-consuming task, becomes a challenge in several RBMT situations due to a lack of reference texts.

### 4.4.2 Performance metric in corpus-based machine translation

The performance metrics for rule-based and data-based machine translation are more or less the same for automatic evaluation. In one of the earliest systems of example-based machine and statistical machine translation, MaTrEx the evaluation was done against a 'gold standard' annotations [115]. Two types of error measurement metrics, WER and PER, measured the distance between reference and candidate translations. Another EBMT system proposed by Almohimeed et al. used the same performance metrics but applied a Leave-One-Out Cross-Validation technique before applying PER and WER [5]. In both, cases accuracy measurement metrics and manual evaluation were left out, thus falling short of being a complete evaluation system.

Lopez-Ludena et al. evaluate the system's performance using both accuracy and error measuring metrics [100]. For accuracy, BLEU and NIST are used, and for error calculation mSER (multiple reference Sign Error Rate) and mPER (multiple Position independent Sign Error Rate) have been added. The system compares three different alternatives of the system, and for comparison, the confidence interval (at 95%) for every BLEU is also presented. This interval is calculated as

$$\pm\Delta = 1.96\sqrt{\frac{\text{BLEU}(100 - \text{BLEU})}{n}}$$

A Spanish to LSE Translation system proposed by San-Segundo et al. evaluates both RBMT and SMT systems in different modules [134]. The evaluation metrics used are SER and BLEU in both cases. These automatic evaluation metrics are capable of indicating RBMT giving better results than SMT. Another evaluation used in this study is the use of confidence measures to inform the user of the translated signs' reliability. Three confidence levels are defined: (1) High Confidence (confidence value higher than 0.5) (2) Medium Confidence (Confidence value between 0.25 and 0.5) (3) Low Confidence (Confidence value less than 0.25). Both the above Spanish translation systems lack manual evaluation.

Though the above systems paved a good way for the use of automatic metrics for evaluation of the translation system, however [132] in a subsequent study, San-Segundo et al. performed an extensive field evaluation of the proposed prototype of Spanish to LSE translation system. The system combined EBMT, SMT, and RBMT translation methods and measured each's performance and integrated, using translation time, SER, PER, BLEU, and NIST metrics. The field evaluation included objective measurements from the system and subjective measures from both Deaf users and government employee questionnaires. The assessment conducted helped to know the actual reviews of deaf people and problems faced by them regarding avatar naturalness.

The systems discussed above majorly produced gloss notations thus relied only on automatic evaluation metrics. In the systems where animation is the final output of the translation, manual evaluation is imperative to evaluate the signed production, especially in the absence of automatic evaluation. In Stein et al.'s approach, the translation is measured using WER and PER metrics, and manual evaluation is conducted using human experts to rate the coherence of the sentence to the avatar output with numbers ranging from 1 (incomprehensive) to 5 (perfect match) [146]. In a Chinese to Taiwanese Sign Language translation system, Su et al.'s proposed approach is compared with the baseline system proposed by Wu et al. using a BLEU metric [148, 162]. The corpus is divided into short and long sentences, and BLEU scores are measured with different n-gram precisions. The second metric, WER, is used on a similar corpus for both proposed and baseline approaches. For manual evaluation, two approaches were followed. Firstly, it adopted mean opinion scores (MOS) method to score translation results for proposed, and baseline approaches. Scores for evaluating the translation were divided into five grades, from 1 for bad and 5 for excellent. The second subjective assessment evaluated the reading comprehension of the ten subjects involved in the evaluation. For this fifth and sixth grade, deaf students were invited as the subject, and a special educator designed 20 questions for testing the reading comprehension of the translated TSL sequences.

Similarly, the MaTrEx model had not synthesized the Irish SL gloss notation into virtual signs. In the following years, this system was extended to include bidirectionality, which added two modules: recognition and SL animation [117]. The translation part of the system was evaluated automatically using BLEU, WER and PER metrics. The animation part of the system was manually evaluated. The evaluation was done in terms of intelligibility (understanding how understandable animation was) and fidelity (assessing how good a translation of the English to the animation was). A scale of 1–4 with qualifying descriptions was used for both metrics. The resultant feedback allowed the evaluator to attribute a completely negative, mostly negative, mostly positive, or completely positive rating to each translation. The evaluation was further concluded with a questionnaire of the general questions regarding translation and technology. This evaluation was done on a web-based format by 4 ISL evaluators provided by the Irish Deaf society.

A Spanish translation system proposed by Lopez-Ludena et al. uses automatic evaluation metrics at different levels [95]. Firstly, the paper presents the BLEU, NIST, mSER, and mPER percentage values of the baseline phrase-based and Statistical Fine State Transducers approach (SFST). Next, a pre-processing module is applied to both methods to reduce the occurrence of errors. The pre-processing module applies tags to the words manually as well as automatically. After applying the pre-processing module to both phrase-based and SFST approaches, automatic evaluation is conducted using the same metrics to see the comparison between automatic and manual tag application and see the effect of the pre-processing module. Lastly, two experts have been involved in manual evaluation to complete the analysis. Both experts evaluated every sentence with one of the three possible scores: 1 (the sentence is well constructed and the meaning is same as original), 0.5 (there are errors but the meaning is same as the original) and 0 (the sentences are not understandable, nor the meaning is same as the original one). Lastly, to show the correlation between automatic metrics and human evaluation, Pearson correlation has been presented. The study has tried to perform an exhaustive assessment of the whole system and kept the avatar presentation of signs and its human evaluation for future research.

A bidirectional EBMT translation process Turkish to Turkish sign language proposed by Selcuk-Simsek et al. highlighted datasets' importance for a fair evaluation [139]. In this system, three datasets were created to observe the characteristics of the system. All three datasets are evaluated separately using BLEU and TER performance metrics. TER metric was also used by Quach et al. in the Vietnamese EBMT system [127].

From the above discussion, it is apparent that BLEU is a critical metric used in collaboration with metrics like TER and WER to measure the system's accuracy and precision.

#### 4.4.3 Performance metrics in neural machine translation (NMT)

The performance metrics used in NMT for automatic evaluation are similar to those used in RBMT or data-driven approaches. In NMT, BLEU is also considered to be one of the most reliable performances metric. ATLASLang NMT is compared with its original data-driven version ATLASLang MTS1 over 73 sentences over a 4-g BLEU score [19, 20].

Apart from BLEU, NMT systems have used other performance metrics. Stoll et al. uses BLEU, ROUGE (ROUGE-L F1) score and WER [147]. This system has been compared with a Gloss2text system on different n-gram granularities [24]. This system also extends its evaluation to evaluate the quality of the generated output. For this, it uses Structured Similarity Index measurement (SSIM), Peak Signal-to-Noise Ratio (PSNR), and Mean Squared Error (MSE) to assess image quality [159]. SSIM is used to measure the perceptual degradation of images and videos in a broadcast by comparing a corrupted image to its original. PSNR and MSE are used to assess the quality of compressed images compared to their original. SIGNGAN approach used the same metrics for measuring the quality of synthesized images [137].

Another NMT study proposed by Jung et al. follows different performance metrics apart from BLEU [73]. In this approach, a complex word reordering technique is formulated. Three MT evaluation metrics used are BLEU, METEOR, and RIBES [35, 69]. METEOR calculates word order similarity using the smallest number of chunks where the system's results can be aligned with a reference sentence [35]. RIBES directly measures the number of reordering events between the system results and reference sentence [69].

The results generated from the score mentioned above were able to show that an NMT system can be created at a low cost and in a less resource environment.

As NMT is a relatively new field compared to RBMT and data-driven, several researchers in this area have not completed the evaluation part of their work. Between manual and automatic evaluation, the latter has been majorly seen in the NMT works. Table 9 catalogs different evaluation measures used in various studies, and Fig. 5 depicts it in a pie chart. In this section state of the art performance metrics have been discussed. BLEU has been used in many studies as it is considered the standard metric for machine translation evaluation [73]. Furthermore, authors have used metrics like RIBES, METEOR to elevate the performance evaluation of the system. As sign synthesis is an integral part of the above studies, metrics like SSIM are used to measure the quality of the synthesized output.

**Table 9** Performance metric's followed by MT studies

| Sr. no | Performance metric | # | Citations |
|---|---|---|---|
| 1 | WER | 11 | [5, 23, 81, 92, 92, 102, 113, 117, 146, 148, 158] |
| 2 | PER | 12 | [5, 23, 95–97, 99, 113, 117, 132, 134, 145, 146] |
| 3 | SER | 7 | [94–97, 99, 132, 134] |
| 4 | TER | 6 | [29, 81, 126, 127, 139, 145] |
| 5 | BLEU | 25 | [2, 20, 73, 79, 81, 89, 94–97, 99, 100, 102, 113, 117, 126, 131, 132, 134, 139, 145, 147, 148, 158, 162] |
| 6 | NIST | 6 | [95–97, 99, 132, 134] |
| 7 | METEOR | 2 | [73, 117] |
| 8 | RIBES | 1 | [73] |
| 9 | Gesture Animation Quality/Gesture Error Rate/Output picture Quality (SSIM, PSNR, MSE) | 3 | [131, 133, 137, 147] |
| 10 | $F1$-score | 1 | [30] |
| 11 | Mean and standard deviation | 1 | [109] |
| 12 | Manual evaluation | 13 | [3, 4, 28–30, 46, 95, 99, 102, 132, 146, 158, 162] |

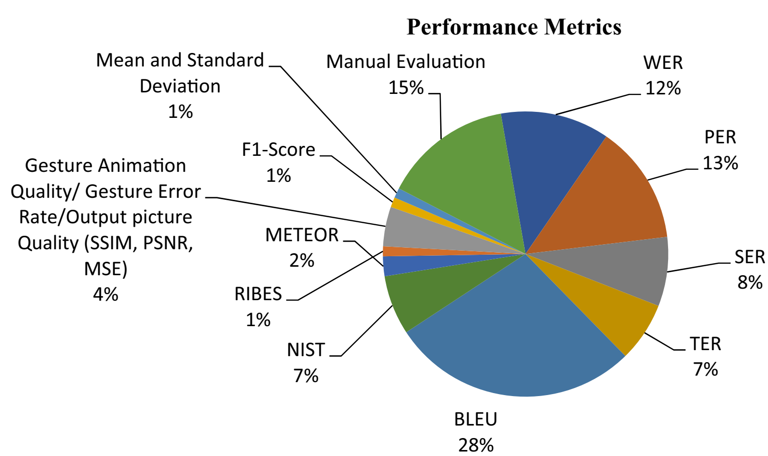



**Fig. 5** Chart depicting performance metrics used in machine translation studies

# 5 Discussion on implications for research and practice

We consider the following avenues as the most promising for future research. These are discussed according to the categories of machine translation, sign synthesis, and a general view.

## 5.1 Rule-based machine translation

A rule-based system is a traditional approach to machine translation. However, due to the lack of bilingual data, rule-based methods produce effective translation systems. The following may be some suggested areas of research.

### 5.1.1 Translating complex sentences

Present research mainly deals with translation of words or simple sentences to sign language, reducing the translation's accuracy. Moreover, this facility is only suitable for to deaf if there are platforms to convert sentences used in everyday conversations, which can help deaf people communicate more effectively and efficiently. Thus, to improve the accuracy and usability of a translation system, complex and compound sentences (sentences with two independent clauses with a related idea) need to be covered.

### 5.1.2 Lack of SL grammar analysis

Academics and researchers worldwide have started researching translating spoken to sign language areas. Several language models have been developed based on linguistic rules understanding; still, many sign languages have not been well-analyzed. In the existing studies, it is noted that apart from ASL and BSL, many sign languages have less formal definitions of grammar rules; thus, they are not able to produce an effective rule-based system. Moreover, lack of grammar also leads to discrepancies in the sign generation module. Therefore, a robust language model and extensive SL rules analysis can be a potential research area in the future.

## 5.2 Corpus-based machine translation

Corpus-based systems do not require handwritten rules and grammar dictionaries; they work on examples, statistical decision theory, and statistical learning. Several existing studies have stated that a corpus-based system works more efficiently in large corpora compared to rule-based systems. Thus corpus-based methods hold significant potentials for future research. Some are listed below:

### 5.2.1 Lack of bilingual corpora in all SL's

Corpus-based machine translation systems work on sentence pairs of spoken and sign languages. A large corpus is essential for data-driven systems (example, statistical, hybrid or neural). Few datasets have been created such as RWTH-Phoenix-14 T, ATIS, DICTA-SIGN, How2Sign, and the latest public DGS corpus [55]. However, the similar size and depth of the corpus do not exist in all sign languages. Henceforth, researchers can focus on creating large datasets for other sign languages.

### 5.2.2 Acquisition of data in multiple SL's

The general studies of SLMT present a single language translation system. Though multilingual datasets like DICTA-SIGN and MultiATIS++ corpus [163] exist, more efforts can be made to bridge communication barriers within different deaf communities. Thus there are plenty of possibilities for future developments in making SL's as any other translating natural language. Henceforth, acquiring data in different sign languages is the area that holds a high future perspective.

## 5.3 Neural machine translation

Neural Networks is a highly utilized concept in various technological developments, but it is a reasonably upcoming concept in SLMT. Certain existing studies have presented that NMT in SLMT gives better results than data-driven or rule-based systems and can produce breakthrough results.

### 5.3.1 Integration with deep learning and AI

Google announced Google Translator's development based on neural machine translation (GMNT). This translator uses an extensive artificial neural network with deep learning capabilities by using millions of examples. An improved version of Google Translator will handle "zero shot translation" which converts directly from one language to another. GNMT supports 101 languages as of September 2020 [172]. For the future, researchers can incorporate AI and deep learning strategies in prevalent NMT systems to achieve similar targets for text to sign translation in any language.

## 5.4 Sign generation

A sign generation system is a crucial component of the SLMT system because it would make more information and services available to the deaf community. They are more fluent in sign language as compared to the corresponding spoken language. Thus sign generation is an essential topic of research.

#### 5.4.1 Naturality of 3D avatar

Prevailing sign synthesis studies reveal that avatar animation is the most suitable sign production form amongst all kinds of sign synthesis. Despite its suitability, avatars have not been successful in being entirely accepted by the deaf community. As an avatar's design requires a considerable amount of hand engineering, its performance remains robotic and under articulated. Non-manual components, facial expressions, eye, and eyebrow movement, can increase the comprehensibility of a 3D avatar. For future studies, researchers can focus on temporal coordination between manual and non-manual components for signing to appear natural.

#### 5.4.2 Alternative methods of sign generation

As the chances of achieving a completely acceptable 3D avatar are slight in the foreseeable future, it is wise to explore other sign synthesis methods. Videos most closely resemble the ground in overall structure and detail. Thus, sign generation capabilities can be analyzed to synthesize meticulous sign videos with signers of acceptable appearance. Furthermore, data processing strategies should be enhanced to focus on SL data's intricate features such as the size of motion and speed.

## 6 Limitations of the review

The main limitation of the present study is that the review is limited to one side of translation, i.e., text to sign language and not the other way around. The data extraction process has been conducted meticulously. A manual search for including all the articles under this field was done. However, some relevant articles published in different languages were not included in this study.

The authors' disagreements arose regarding inclusion and exclusion of some studies that were resolved with discussions, and a joint agreement was achieved. The experience of one author in conducting a systematic review was used notably at each stage. Study categories were decided based upon the available literature in the collected studies.

## 7 Conclusion

Text to Sign Language Translation has been a widely researched area amongst various communities worldwide working for the betterment of deaf societies. A proof of this is the number of publications which come out every year. We have conducted a literature review to review the number of techniques being used in this field. From more than 200 papers, we have selected 151 studies published in approximately 40 conferences and workshops and 30 reputed journals which contributed directly to this field. We classified the papers according to different types of machine translation systems, sign language generation methods, and evaluation metrics used.

After analyzing the selected papers, we have noticed that the number of publications has been consistent in the last two decades. Although the balance of techniques has changed from conventional to contemporary styles, traditional techniques are still noticed. We have seen that in the nineties and early 2000, RBMT held the monopoly over the machine translation field. The need for a larger dataset increased the path for data-driven machine translation was paved. Another consideration is that many papers have focussed on creating large bilingual corpora, which is an essential requirement of data-driven systems.

Several research efforts conducted do not directly deal with the translation process but focus on the pre-processing of data and sign synthesis. Sign synthesis has been the main focus of many researchers because it is relevant to the deaf. Several kinds of sign generations have been reported. Though a significant amount of research is done in the sign synthesis field, still a level of completion and satisfaction for the deaf has not been achieved.

Evaluation is an essential aspect in SLMT as the feedback of the deaf is the best source to understand the use, effectiveness, and limitation of proposed projects. We have reported both manual and automatic evaluations and their respective performance metrics used in all the studies.

Lastly, we have identified several gaps in sign language translation and generation from the complete analysis. RBMT requires strong linguistic knowledge, which limits the translation system to a particular language. The amount of bilingual data is scarce in many languages, making it hard for data-driven techniques to produce accurate results. Even after two decades of research, the sign generation module lacks being natural, flexible, and comfortable as required by the deaf.

Artificial Neural Networks is the latest and upcoming technology that has shown promising results in other fields. Based upon it, neural machine translation can achieve the much awaited breakthrough in Text-to-Sign language translation in the coming future.

## Appendix A: A quality assessment form

### Screening question

| | |
|---|---|
| Does the research paper refer to Sign language machine translation? | ☐ Yes |
| Consider: | ☐ No |
| The paper includes the study of machine translation for translation between spoken language and sign language. All the research papers included are related to the field of sign language | |

If the reply is positive for the above question, we proceed to the following screening question.

### Screening question

| | |
|---|---|
| Is the focus of research paper machine translation techniques for text to sign language conversion? | ☐ Yes |
| Consider: | ☐ No |
| Is the study's main focus text to sign language machine translation or not? | |
| Did the study fit in any of the sub-areas of machine translations categorized? | |

If the study's primary focus is text to SLMT, then the following detailed questions are considered.

### Detailed questions

| | |
|---|---|
| Findings | |
| Is there a clear statement of findings? | ☐ Yes |
| Consider: | ☐ No |
| Did the study mention the machine translation technique for SL translation? | |
| What is the corresponding sign generation method? | |
| Was the data reported sufficient for comparative analysis? | ☐ Yes |
| Consider: | ☐ No |
| Are the necessary parameters for comparison discussed? | |
| Were the evaluation methods discussed? | ☐ Yes |
| Consider: | ☐ No |
| Were the necessary performance metrics for manual and automatic evaluation discussed? | |

**Funding** No funding for this manuscript.

**Availability of data and material** Not Applicable.

## Declarations

**Conflict of interest** No conflict of interest.

## References

1. Boudlal, A., Lakhoujaja, A.M.: Alkhalil Morpho Sys1: a morphosyntactic analysis system for Arabic texts. In: International Arab Conference on Information Technology, pp. 1–6 (2010)
2. Achraf, O., Jemni, M.: Designing high accuracy statistical machine translation for sign language using parallel corpus: case study English and American sign language. J. Inf. Technol. Res. **12**(2), 134–158 (2019). https://doi.org/10.4018/JITR.2019040108
3. Al-barahamtoshy, O.H., Al-barhamtoshy, H.M.: Arabic text-to-sign (ArTTS ) model from automatic SR system. In: Proceedings of 3rd International Conference on Arab Computational Linguistics, pp. 304–311 (2017). https://doi.org/10.1016/j.procs.2017.10.122
4. Almeida, I.: Exploring challenges in avatar-based translation from European Portuguese to Portuguese sign language. Master's Thesis. Instituto Superior Técnico (2014)
5. Almohimeed, A. et al.: Arabic text to Arabic sign language translation system for the deaf and hearing-impaired community. In: Proceedings of the second workshop on speech and language processing for assistive technologies, pp 101–109 (2011)
6. Anuja, K. et al.: Design and development of a frame based MT system for English-to-ISL. In: Proceedings of the World congress on nature and biologically inspired computing, pp. 1382–1387 (2009). https://doi.org/10.1109/NABIC.2009.5393721
7. Araújo, T.M.U.D., et al.: An approach to generate and embed sign language video tracks into multimedia contents. Inf. Sci. (Ny) **281**, 762–780 (2014). https://doi.org/10.1016/j.ins.2014.04.008
8. Badler, N.I., et al.: Animation control for real-time virtual humans. Commun. ACM. **42**(8), 64–73 (1999). https://doi.org/10.1145/310930.310975
9. Baldassarri, S. et al.: Automatic translation system to Spanish sign language with a virtual interpreter. In: Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction, pp. 196–199 (2009). https://doi.org/10.1007/978-3-642-03655-2_23
10. Baldassarri, S., et al.: Maxine: a platform for embodied animated agents. Comput. Graph. **32**(4), 430–437 (2008). https://doi.org/10.1016/j.cag.2008.04.006
11. Baldassarri, S., Royo-Santas, F.: An automatic rule-based translation system to Spanish sign language (LSE). In: New Trends in Human–Computer Interaction, pp. 1–11. https://doi.org/10.1007/978-1-84882-352-5_1
12. Bangham, J.A. et al.: Signing for the deaf using virtual humans. In: IEE Seminar on Speech and Language Processing for the Disabled and Elderly People, pp. 1–5, London, UK (2000). https://doi.org/10.1049/ic:20000134
13. Bangham, J.A. et al.: Virtual signing: capture, animation , storage and transmission—an overview of the ViSiCAST project. In: IEEE Seminar on Speech and Language Processing for Disabled and Elderly People, pp. 6/1–6/7 IEEE Press (2000). https://doi.org/10.1049/ic:20000136
14. Bishop, C.M.: Mixture density networks. Technical report (1994)
15. Boulares, M., et al.: Toward an example-based machine translation from written text to ASL using virtual agent animation. Int. J. Comput. Sci. Issues **9**, 1 (2012)

16. Bowden, R. et al.: Learning to recognise dynamic visual content from broadcast footage. https://cvssp.org/projects/dynavis/index.html. Last Accessed 2021/04/04
17. Braffort, A., et al.: KAZOO: a sign language generation platform based on production rules. Univers. Access Inf. Soc. **15**(4), 541–550 (2016). https://doi.org/10.1007/s10209-015-0415-2
18. Braffort, A. et al.: Virtual signer coarticulation in octopus, a sign language generation platform. In: Proceedings of the 9th International Gesture Workshop, Gesture in Embodied Communication and Human–Computer Interaction, pp. 29–32 (2011)
19. Brour, M., Benabbou, A.: ATLASLang MTS 1: Arabic text language into Arabic sign language machine translation system. In: 2nd International Conference on Intelligent Computing in Data Sciences, pp. 236–245 (2019). https://doi.org/10.1016/j.procs.2019.01.066
20. Brour, M., Benabbou, A.: ATLASLang NMT: Arabic text language into Arabic sign language neural machine translation. J. King Saud Univ. Comput. Inf. Sci. (2019). https://doi.org/10.1016/j.jksuci.2019.07.006
21. Bungeroth, J. et al.: The ATIS sign language corpus. In: Proceedings of the 6th International Conference on Language Resources and Evaluation, pp. 2943–2946 (2008)
22. Bungeroth, J., Ney, H.: Statistical sign language translation. In: Workshop on Representation and Processing of Sign Languages, pp. 105–108 (2004)
23. Buz, B., Gungor, T.: Developing a statistical Turkish sign language translation system for primary school students. In: IEEE International Symposium on Innovations in Intelligent SysTems and Applications, pp. 1–6 (2019). https://doi.org/10.1109/INISTA.2019.8778246
24. Camgoz, N.C. et al.: Neural sign language translation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 7784–7793 (2018). https://doi.org/10.1109/CVPR.2018.00812
25. Chan, C. et al.: Everybody dance now. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 5933–5942 (2019)
26. Chung, J. et al.: Empirical evaluation of gated recurrent neural networks on sequence modeling. Presented at the (2014)
27. Coetzee, L. et al.: The national accessibility portal: an accessible information sharing portal for the South African disability sector. In: Proceedings of the International Cross-Disciplinary Conference on Web Accessibility, pp. 44–53. Banff, Canada (2007). https://doi.org/10.1145/1243441.1243456
28. Cox, S., et al.: The development and evaluation of a speech-to-sign translation system to assist transactions. J. Hum. Comput. Interact. **16**(2), 141–161 (2003). https://doi.org/10.1207/S15327590IJHC1602
29. Da, Q.L., et al.: Converting the vietnamese television news into 3D sign language animations for the deaf. In: Duong, T., Vo, N.S. (eds.) Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 257. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-05873-9_13
30. Dangsaart, S., et al.: Intelligent Thai text—Thai sign translation for language learning. Comput. Educ. **51**(3), 1125–1141 (2008). https://doi.org/10.1016/j.compedu.2007.11.008
31. Dangsaart, S., Cercone, N.: Bridging the gap: Thai–Thai sign machine translation. In: Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics, pp. 191–199 (2007)
32. Dasgupta, T., Basu, A.: Prototype machine translation system from text-to-Indian sign language. In: Proceedings of the 13th International conference on Intelligent User Interfaces, pp. 313–316 (2008). https://doi.org/10.1145/1378773.1378818
33. Davidson, M.J.: PAULA: a computer-based sign language tutor for hearing adults. In: Intelligent Tutoring Systems Workshop on Teaching with Robots, Agents, and Natural Language Processing, pp. 66–72 (2006)
34. Delorme, M. et al.: Thumb modelling for the generation of sign language. In: Proceedings of the 9th International Conference on Gesture and Sign Language in Human–Computer Interaction and Embodied Communication, pp. 151–160 (2012). https://doi.org/10.1007/978-3-642-34182-3_14
35. Denkowski, M., Lavie, A.: Meteor universal: language specific translation evaluation for any target language. In: Proceedings of 9th Workshop on Statistical Machine Translation, pp. 376–380 (2014). https://doi.org/10.3115/v1/W14-3348
36. Doddington, G.: Automatic evaluation of machine translation quality using N-gram co-occurrence statistics. In: Proceedings of the 2nd International Conference on Human Language Technology Research, pp. 138–145 (2002)
37. Duarte, A.: Cross-modal neural sign language translation. In: In Proceedings of the 27th ACM International Conference on Multimedia, pp. 1650–1654 (2019). https://doi.org/10.1145/3343031.3352587
38. Ebling, S. et al.: SMILE Swiss German sign language dataset. In: Proceedings of the International Conference on Language Resources and Evaluation, pp. 19–25 (2018)
39. Ebling, S., Huenerfauth, M.: Bridging the gap between sign language machine translation and sign language animation using sequence classification. In: Proceedings of the 6th Workshop on Speech and Language Processing for Assistive Technologies, pp. 2–9 (2015). https://doi.org/10.18653/v1/W15-5102
40. Efthimiou, E. et al.: DICTA-SIGN: sign language recognition, generation, and modelling: a research effort with applications in deaf communication. In: Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, pp. 80–84 (2009). https://doi.org/10.1007/978-3-642-02707-9_3
41. Efthimiou, E., et al.: Feature-based natural language processing for GSL synthesis. Sign Lang. Linguist. **10**(1), 1–21 (2007). https://doi.org/10.1075/sll.10.1.03eft
42. Efthimiou, E., Dimou, S.F.A.: From grammar-based MT to post-processed SL representations. Univers. Access Inf. Soc. **15**(4), 499–511 (2016). https://doi.org/10.1007/s10209-015-0414-3
43. Elliott, R. et al.: The development of language processing support for the ViSiCAST project. In: Proceedings of the 4th International ACM Conference on Assistive Technologies, pp. 101–108 (2000). https://doi.org/10.1145/354324.354349
44. Eryiğit, C., et al.: Building machine-readable knowledge representations for Turkish sign language generation. Knowl.-Based Syst. **108**, 179–194 (2016). https://doi.org/10.1016/j.knosys.2016.04.014
45. Eryiğit, G.: ITU Turkish NLP web service. In: Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics, pp. 1–4 (2014). https://doi.org/10.3115/v1/E14-2001
46. Evita Fotinea, S., et al.: A knowledge-based sign synthesis architecture. Univers. Access Inf. Soc. **6**(4), 405–418 (2008). https://doi.org/10.1007/s10209-007-0094-8
47. Fellbaum, C. (ed.): WordNet: An Electronic Lexical Database. MIT Press, Massachusetts (1998)
48. Filhol, M., et al.: A rule triggering system for automatic text-to-sign translation. Univers. Access Inf. Soc. **15**(4), 487–498 (2016). https://doi.org/10.1007/s10209-015-0413-4
49. Filhol, M.: Combining two synchronisation methods in a linguistic model to describe sign language. In: Proceedings of the 9th International Conference on Gesture and Sign Language in Human–Computer Interaction and Embodied Communication,

pp. 194–203 (2011). https://doi.org/10.1007/978-3-642-34182-3_18
50. Glauert, J., et al.: Linguistic modelling and language-processing technologies for Avatar-based sign language presentation. Univers. Access Inf. Soc. **6**(4), 375–391 (2008). https://doi.org/10.1007/s10209-007-0102-z
51. Glauert, J.R.W., et al.: VANESSA—a system for communication between deaf and hearing people. Technol. Disabil. **18**(4), 207–216 (2006). https://doi.org/10.3233/TAD-2006-18408
52. Gough, N.: Example-based machine translation using the marker hypothesis. PhD thesis. Dublin City University (2005)
53. Grieve-Smith, A.B.: English to American sign language machine translation of weather reports. In: Proceedings of the Second High Desert Student Conference in Linguistics. High Desert Linguistics Society, pp. 23–30 (1999)
54. Grieve-Smith, A.B.: SignSynth: A sign language synthesis application using Web3D and Perl. In: Gesture and sign language based human–computer interaction, pp. 134–145. London, UK (2001). https://doi.org/10.1007/3-540-47873-6_14
55. Hanke, T. et al.: Extending the public DGS corpus in size and depth. In: Proceedings of the 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 (2020)
56. Hanke, T.: HamNoSys—Representing sign language data in language resources and language processing contexts. In: Proceedings of the LREC workshop on the representation and processing of sign languages, pp. 1–6 (2004)
57. Hanke, T., Popescu, H.: eSIGN deliverable D2.3: intelligent sign editor (2003)
58. Heloir, A., Kipp, M.: Real-time animation of interactive agents: specification and realization. Appl. Artif. Intell. **24**(6), 510–529 (2010). https://doi.org/10.1080/08839514.2010.492161
59. Hogan, C., Frederking, R.: An evaluation of the multi-engine MT architecture. In: Proceedings of the Conference of the Association for Machine Translation in the Americas, pp. 113–123 (1998). https://doi.org/10.1007/3-540-49478-2_11
60. Huang, Z., Eli, A.: STEP: a scripting language for embodied agents. In: Proceedings of the Workshop of Lifelike Animated Agents, pp. 1–6 (2002)
61. Huenerfauth, M.: A linguistically motivated model for speed and pausing in animations of American sign language. ACM Trans. Access. Comput. **2**, 2 (2009). https://doi.org/10.1145/1530064.1530067
62. Huenerfauth, M.: A multi-path architecture for machine translation of English text into American sign language animation. In: Proceedings of the Student Workshop at the Human Language Technology Conference/North American Chapter of the Association for Computational Linguistics, pp. 25–30 (2004). https://doi.org/10.3115/1614038.1614043
63. Huenerfauth, M.: American sign language generation: multimodal NLG with multiple linguistic channels. In: Proceedings of the ACL Student Research Workshop, pp. 37–42 (2005). https://doi.org/10.5555/1628960.1628968
64. Huenerfauth, M.: An accessibility motivation for an English-to-ASL machine translation system (2004)
65. Huenerfauth, M.: Generating American sign language classifier predicates for English-to-ASL machine translation. Ph.D thesis. University of Pennsylvania (2006)
66. Huenerfauth, M.: Spatial representation of classifier predicates for machine translation into American sign language. In: Proceedings of the Workshop on the Representation and Processing of Signed Languages, 4th International Conference on Language Resources and Evaluation, pp. 24–31 (2004)
67. Huenerfauth, M., Lu, P.: Effect of spatial reference and verb inflection on the usability of sign language animations. Univers. Access Inf. Soc. **11**(2), 169–184 (2012). https://doi.org/10.1007/s10209-011-0247-7
68. Huenerfauth, M.P.: A survey and critique of American sign language natural language generation and machine translation systems. Technical report (2003)
69. Isozaki, H. et al.: Automatic evaluation of translation quality for distant language pairs. In: Proceedings of Conference on Empirical Methods and Natural Language Processing, pp. 944–952 (2010). https://doi.org/10.5555/1870658.1870750
70. Jemni, M. et al.: A Web-based tool to create online courses for deaf pupils. In: Proceedings of the International Conference on Interactive Mobile and Computer Aided Learning, pp. 1–8. Amman, Jordan (2007)
71. Jemni, M., Elghoul, O.: A system to make signs using collaborative approach. In: International Conference on Computers for Handicapped Persons. Lecture Notes in Computer Science, pp. 670–677 (2008). https://doi.org/10.1007/978-3-540-70540-6_96
72. Jemni, M., Elghoul, O.: Towards Web-based automatic interpretation of written text to sign language. In: Proceedings of the 1st International Conference on ICT & Accessibility, pp. 43–48. (2008).
73. Jung, H.Y., et al.: Word reordering for translation into Korean sign language using syntactically-guided classification. ACM Trans. Asian Low-Resource Lang. Inf. Process. **19**(2), 1–20 (2019). https://doi.org/10.1145/3357612
74. Kanis, J. et al.: Czech-sign speech corpus for semantic based machine translation. In: Sojka, P., Kopeček, I., Pala, K. (eds.) Proceedings of 9th International Conference on Text, Speech and Dialogue, pp. 613–620 (2006). https://doi.org/10.1007/11846406_77
75. Kanis, J., Müller, L.: Automatic Czech—sign speech translation. In: Proceedings of 10th International Conference on Text, Speech and Dialogue, pp. 488–495 (2007). https://doi.org/10.1007/978-3-540-74628-7_63
76. Kar, P. et al.: INGIT: Limited domain formulaic translation from Hindi strings to Indian sign language. In: International Conference on Natural Language Processing (2007)
77. Karpouzis, K., Caridakis, G.: Educational resources and implementation of a Greek sign language synthesis architecture. Comput. Educ. **49**(1), 54–74 (2007). https://doi.org/10.1016/j.compedu.2005.06.004
78. Katyana, Q.: Google's neural network learns to translate languages it hasn't been trained on. https://www.theregister.co.uk/2016/11/17/googles_neural_net_translates_languages_not_trained_on/, Last accessed 2020/09/23
79. Kayahan, D., Gungor, T.: A hybrid translation system from Turkish spoken language to Turkish sign language. In: IEEE international symposium on innovations in intelligent systems and applications, pp. 1–6 (2019). https://doi.org/10.1109/INISTA.2019.8778347
80. Kennaway, R.: Synthetic animation of deaf signing gestures. In: 4th International Workshop on Gesture and Sign Language Based Human–Computer Interaction, pp. 146–157 (2002). https://doi.org/10.1007/3-540-47873-6_15
81. Khan, N.S., et al.: A novel natural language processing (NLP)—based machine translation model for English to Pakistan sign language translation. Cognit. Comput. **12**, 748–765 (2020). https://doi.org/10.1007/s12559-020-09731-7
82. Kipp, M. et al.: Sign language avatars: animation and comprehensibility. In: Proceedings of the 10th International Conference on Intelligent Virtual Agents, pp. 113–126 (2011). https://doi.org/10.1007/978-3-642-23974-8

83. Kitchenham, B., Charters, S.: Guidelines for performing systematic literature reviews in software engineering. Technical report EBSE-2007-01. (2007)
84. Koehn, P. et al.: Moses: open source toolkit for statistical machine translation. In: Companion Volume to the Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, pp. 177–180 (2007)
85. Koehn, P. et al.: Statistical phrase-based translation. In: Proceedings of the Human Language Technology and North American Association for Computational Linguistics Conference, pp. 48–54 (2003)
86. Koleli, E.: A new Greek part-of-speech tagger, based on a maximum entropy classifier. Master's thesis. Athens University of Economics and Business (2011)
87. Kouremenos, D., et al.: A novel rule based machine translation scheme from Greek to Greek sign language: production of different types of large corpora and language models evaluation. Comput. Speech Lang. **51**, 110–135 (2018). https://doi.org/10.1016/j.csl.2018.04.001
88. Kouremenos, D., et al.: A prototype Greek text to Greek sign language conversion system. Behav. Inf. Technol. **29**(5), 467–481 (2010). https://doi.org/10.1080/01449290903420192
89. Kouremenos, D. et al.: Statistical machine translation for Greek to Greek sign language using parallel corpora produced via rule-based machine translation. In: IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), pp. 1–15 (2018)
90. Kovar, L. et al.: Motion graphs. In: Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, pp. 473–482 (2002). https://doi.org/10.1145/566570.566605
91. Krnoul, Z. et al.: 3D symbol base translation and synthesis of Czech sign speech. In: Proceedings of the 11th International Conference on Speech and Computer, pp. 530–535 (2006)
92. Krňoul, Z., Železný, M.: Translation and conversion for Czech sign speech synthesis. Lect. Notes Comput. Sci. **4629**, 524–531 (2007). https://doi.org/10.1007/978-3-540-74628-7_68
93. Le, H.P. et al.: A hybrid approach to word segmentation of Vietnamese texts. In: Proceedings of 2nd International Conference on Language and Automata Theory and Applications, pp. 240–249 (2008). https://doi.org/10.1007/978-3-540-88282-4_23
94. López-ludeña, V., et al.: Translating bus information into sign language for deaf people. Eng. Appl. Artif. Intell. **32**, 258–269 (2014). https://doi.org/10.1016/j.engappai.2014.02.006
95. López-Ludeña, V., et al.: Automatic categorization for improving Spanish into Spanish Sign Language machine translation. Comput. Speech Lang. **26**(3), 149–167 (2012). https://doi.org/10.1016/j.csl.2011.09.003
96. López-Ludeña, V. et al.: Factored translation models for improving a speech into sign language translation system. In: Proceedings of the Conference of the International Speech Communication Association, pp. 1605–1608 (2011)
97. López-Ludeña, V., et al.: Increasing adaptability of a speech into sign language translation system. Expert Syst. Appl. **40**(4), 1312–1322 (2013). https://doi.org/10.1016/j.eswa.2012.08.059
98. López-Ludeña, V., et al.: Methodology for developing an advanced communications system for the deaf in a new domain. Knowl.-Based Syst. **56**, 240–252 (2014). https://doi.org/10.1016/j.knosys.2013.11.017
99. Lopez Ludeña, V. et al.: Methodology for developing a speech into sign language translation system in a new semantic domain. In: Proceedings of the Conference on IberSPEECH, pp. 193–203. Madrid, Spain (2012)
100. Ludeña, V.L., San-segundo, R.: Statistical methods for improving Spanish into Spanish sign language translation. In: Proceedings of the 15th Mexican International Conference on Artificial Intelligence, pp. 1–11 (2016)
101. Luong, M.T. et al.: Effective approaches to attention-based neural machine translation. In: Conference on Empirical Methods in Natural Language Processing, pp. 1412–1421 (2015). https://doi.org/10.18653/v1/d15-1166
102. Luqman, H., Mahmoud, S.A.: Automatic translation of Arabic text-to-Arabic sign language. Univers. Access Inf. Soc. **18**(4), 939–951 (2019). https://doi.org/10.1007/s10209-018-0622-8
103. Manzano, D.M.: English to ASL translator for SPEECH2SIGNS. (2018)
104. Marshall, I., Safar, E.: Extraction of semantic representations from syntactic CMU link grammar linkages. In: Recent Advances in Natural Language Processing, pp. 154–159 (2001)
105. Marshall, I., Safar, E.: Grammar Development for Sign Language Avatar-Based Synthesis . In: Proceedings of the 11th International Conference on Human Computer Interaction. (2005).
106. Marshall, I., Safar, E.: Sign language generation in an ALE HPSG. In: Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar, pp. 189–201 (2004)
107. Marshall, I., Sáfár, É.: A prototype text to British sign language (BSL) translation system. In: 41st Annual Meeting of the Association for Computational Linguistics, pp. 113–116 (2003). https://doi.org/10.3115/1075178.1075194
108. Mauser, A., Ney, H.: Automatic evaluation measures for statistical machine translation system optimization. In: Proceedings of the 6th International Conference on Language Resources and Evaluation, pp. 28–30
109. Mazzei, A. et al.: Deep natural language processing for Italian Sign Language translation. In: Proceedings of the 13th Conference of the Italian Association for Artificial Intelligence, pp. 193–204 (2013). https://doi.org/10.1007/978-3-319-03524-6_17
110. Mishra, G.S., et al.: Word based statistical machine translation from English text to Indian sign language. ARPN J. Eng. Appl. Sci. **12**(2), 481–489 (2017)
111. Mitkov, R. (ed.): The Oxford Handbook of Computational Linguistics. Oxford University Press, Oxford (2005)
112. Miyazaki, T. et al.: Proper name machine translation from Japanese to Japanese sign language. In: Language Technology for Closely Related Languages and Language Variants, pp. 67–75 (2014). https://doi.org/10.3115/v1/w14-4209
113. Morrissey, S.: Assistive technology for deaf people: Translating into and animating Irish sign language. In: Proceedings of the 12th International Conference on Computers Helping People with Special Needs, pp. 8–14 (2008)
114. Morrissey, S., Way, A.: An example-based approach to translating sign language. In: Workshop example-based machine translation (MT X-05), pp. 109–116 (2005)
115. Morrissey, S., Way, A.: Joining hands: developing a sign language machine translation system with and for the deaf community. In: Proceedings of the Conference and Workshop on Assistive Technologies for People with Vision & Hearing Impairments, pp. 1–6 (2007)
116. Morrissey, S., Way, A.: Lost in translation: the problems of using mainstream MT Evaluation metrics for sign language translation. In: Proceedings of the 5th SALTMIL Workshop on Minority Languages at Language Resources and Evaluation Conference, pp. 91–98 (2006)
117. Morrissey, S., Way, A.: Manual labour: tackling machine translation for sign languages. Mach. Transl. **27**(1), 25–64 (2013). https://doi.org/10.1007/s10590-012-9133-1
118. Nagao, M.: Framework of a mechanical translation between Japanese and English by analogy principle. In: Proceedings of the International NATO symposium on artificial and human intelligence, pp. 173–180 (1984). https://doi.org/10.7551/mitpress/5779.003.0038

119. Needleman, S.B., Wunsch, C.D.: A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. **48**(3), 443–453 (1970). https://doi.org/10.1016/0022-2836(70)90057-4
120. Nguyen, T.B.D. et al.: A rule based method for text shortening in Vietnamese sign language translation. In: Information Systems Design and Intelligent Applications. Advances in Intelligent Systems and Computing (2018). https://doi.org/10.1007/978-981-10-7512-4_65
121. Nießen, S. et al.: An evaluation tool for machine translation: fast evaluation for MT research. In: Proceedings of 2nd International Conference on Language Resources and Evaluation, pp. 39–45 (2000)
122. Othman, A., Tmar, Z.: English-ASL gloss parallel corpus 2012: ASLG-PC12. In: 5th Workshop on the Representation and Processing of Sign Languages (2012)
123. Otoom, M., Alzubaidi, M.A.: Ambient intelligence framework for real-time speech-to-sign translation. Assist. Technol. **30**(3), 119–132 (2018). https://doi.org/10.1080/10400435.2016.1268218
124. Papageorgiou, H. et al.: A unified POS tagging architecture and its application to Greek. In: Proceedings of the 2nd Language Resources and Evaluation Conference, pp. 1455–1462 (2000)
125. Papineni, K. et al.: BLEU: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pp. 311–318 (2002). https://doi.org/10.3115/1073083.1073135
126. Porta, J., et al.: A rule-based translation from written Spanish to Spanish sign language glosses. Comput. Speech Lang. **28**(3), 788–811 (2014). https://doi.org/10.1016/j.csl.2013.10.003
127. Quach, L., Nguyen, C.-N.: Conversion of the Vietnamese grammar into sign language structure using the example-based machine translation algorithm. In: International Conference on Advanced Technologies for Communications, pp. 27–31 (2018). https://doi.org/10.1109/ATC.2018.8587584
128. Safar, E., Glauert, J.: Computer modelling. In: Pfau, R., et al. (eds.) Sign Language, pp. 1075–1102. De Gruyter Mouton, Berlin (2012). https://doi.org/10.1515/9783110261325.1075
129. Safar, E., Marshall, I.: The architecture of an English-text-to-sign-languages translation system. In: Recent Advances in Natural Language Processing, pp. 223–228. Bulgaria (2001)
130. Sáfár, É., Marshall, I.: Sign language translation via DRT and HPSG. In: Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics, pp. 58–68 (2002). https://doi.org/10.1007/3-540-45715-1_5
131. San-Segundo, R. et al.: A Spanish speech to sign language translation system for assisting deaf-mute people. In: Proceedings of the 9th International Conference on Spoken Language Processing, pp. 1399–1402 (2006)
132. San-Segundo, R., et al.: Design, development and field evaluation of a Spanish into sign language translation system. Pattern Anal. Appl. **15**(2), 203–224 (2012). https://doi.org/10.1007/s10044-011-0243-9
133. San-Segundo, R., et al.: Proposing a speech to gesture translation architecture for Spanish deaf people. J. Vis. Lang. Comput. **19**(5), 523–538 (2008). https://doi.org/10.1016/j.jvlc.2007.06.002
134. San-Segundo, R., et al.: Speech to sign language translation system for Spanish. Speech Commun. **50**(11–12), 1009–1020 (2008). https://doi.org/10.1016/j.specom.2008.02.001
135. Sandler, W., Lillo-Martin, D.: Sign Language and Linguistic Universals. J. Linguist. **42**(3), 738–742 (2006). https://doi.org/10.1017/CBO9781139163910
136. Saunders, B. et al.: Adversarial training for multi-channel sign language production. In: Proceedings of the British Machine Vision Conference (2020)
137. Saunders, B. et al.: Everybody sign now: translating spoken language to photo realistic sign language video (2020)
138. Saunders, B. et al.: Progressive transformers for end-to-end sign language production. In: Proceedings of European Conference on Computer Vision, pp. 687–705 (2020). https://doi.org/10.1007/978-3-030-58621-8_40
139. Selcuk-Simsek, M., Cicekli, I.: Bidirectional machine translation between Turkish and Turkish sign language: a data-driven approach. Int. J. Nat. Lang. Comput. **6**(3), 33–46 (2017). https://doi.org/10.5121/ijnlc.2017.6303
140. Shieber, S.M., Yves, S.: Synchronous tree-adjoining grammars. In: Proceedings of the 13th International Conference on Computational Linguistics, pp. 253–258 (1990). https://doi.org/10.3115/991146.991191
141. Sleator, D.D., Temperley, D.: Parsing English with a link grammar. Technical report CMU-CS-91-196 (1991)
142. Snover, M. et al.: Study of translation edit rate with targeted human annotation. In: Proceedings of the 7th Conference of the Association for Machine Translation in the Americas, pp. 223–231. Cambridge, MA (2006)
143. Souteh, Y., Bouzoubaa, K.: SAFAR platform and its morphological layer. In: Proceedings of the 11th Conference on Language Engineering, pp. 14–15 (2011)
144. Speers, d' A.L.: Representation of American sign language for machine translation. Ph.D thesis. Georgetown University (2001)
145. Stein, D., et al.: Analysis, preparation, and optimization of statistical sign language machine translation. Mach. Transl. **26**(4), 325–357 (2012). https://doi.org/10.1007/s10590-012-9125-1
146. Stein, D. et al.: Morpho-syntax based statistical methods for automatic sign language translation. In: Proceedings of the 11th Annual Conference of the European Association for Machine Translation, pp. 169–177. Oslo, Norway (2006)
147. Stoll, S., et al.: Text2Sign: towards sign language production using neural machine translation and generative adversarial networks. Int. J. Comput. Vis. **128**, 891–908 (2020). https://doi.org/10.1007/s11263-019-01281-2
148. Su, H.Y., Wu, C.H.: Improving structural statistical machine translation for sign language with small corpus using thematic role templates as translation memory. IEEE Trans. Audio Speech Lang. Process. **17**(7), 1305–1315 (2009). https://doi.org/10.1109/TASL.2009.2016234
149. Suszczanska, N. et al.: Translating Polish texts into sign language in the TGT system. In: 20th IASTED International Multi-Conference on Applied Informatics, pp. 282–287 (2002)
150. Sutton, C., McCallum, A.: An introduction to conditional random fields By Charles Sutton and Andrew McCallum. Found. Trends Mach. Learn. **4**(4), 267–373 (2012). https://doi.org/10.1561/2200000013
151. Tillmann, C. et al.: Accelerated DP based search for statistical translation. In: Proceedings of the 5th European Conference on Speech Communication and Technology, pp. 2667–2670 (1997)
152. Tokuda, M., Okumura, M.: Towards automatic translation from Japanese into Japanese sign language. Assist. Technol. Artif. Intell. Robot. User Interfaces Nat. Lang. Process. **1458**, 97–108 (1998). https://doi.org/10.1007/bfb0055973
153. Toutanova, K. et al.: Feature-rich part-of-speech tagging with a cyclic dependency network. In: Conference of the North American Chapter of the Association for Computational Linguistics & Human Language Technologies on Human Language Technologies, pp. 173–180 (2003)
154. Veale, T., et al.: The challenges of cross-modal translation: English-to-sign-language translation in the Zardoz system. Mach. Transl. **13**(1), 81–106 (1998). https://doi.org/10.1023/A:1008014420317
155. Veale, T., Collins, B.: Space, metaphor and schematization in sign: sign language translation in the ZARDOZ system.

In: Proceedings of the 2nd Conference of the Association for Machine Translation in the Americas, pp. 168–179 (1996)
156. Veale, T., Conway, A.: Cross modal comprehension in ZARDOZ an English to sign-language translation system. In: 4th International Workshop on Natural Language Generation, pp. 67–72 (1994). https://doi.org/10.3115/1641417.1641450
157. Ventura, L. et al.: Can everybody sign now? Exploring sign language video generation from 2D poses. In: Sign Language Recognition, Translation and Production Workshop (2020)
158. Veríssimo, V.M., et al.: Towards an open platform for machine translation of spoken languages into sign languages. Mach. Transl. **33**(4), 315–348 (2019). https://doi.org/10.1007/s10590-019-09238-5
159. Wang, Z., et al.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861
160. Waterman, M., Smith, T.: Identification of common molecular subsequences. J. Mol. Biol. **147**(1), 195–197 (1981). https://doi.org/10.1016/0022-2836(81)90087-5
161. Wray, A., et al.: A formulaic approach to translation at the post office: reading the signs. Lang. Commun. **24**, 59–75 (2004). https://doi.org/10.1016/j.langcom.2003.08.001
162. Wu, C.H., et al.: Transfer-based statistical translation of Taiwanese sign language using PCFG. ACM Trans. Asian Lang. Inf. Process. **6**(1), 1–18 (2007). https://doi.org/10.1145/1227850.1227851
163. Xu, W. et al.: End-to-end slot alignment and recognition for cross-lingual NLU. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 5052–5063 (2020)
164. Zhao, L. et al.: A machine translation system from English to American sign language. In: Proceedings of the 4th Conference of the Association of Machine Translation, pp. 293–300 (2000). https://doi.org/10.1007/3-540-39965-8_6
165. Van Zijl, L.: South African sign language machine translation project. In: Proceedings of the 8th International Conference on Computers and Accessibility, pp. 233–234 (2006). https://doi.org/10.1145/1168987.1169031
166. Van Zijl, L., Combrink, A.: The South African sign language machine translation project: issues on non-manual sign generation. In: Proceedings of the Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries, pp. 127–134 (2006). https://doi.org/10.1145/1216262.1216276
167. Van Zijl, L., Olivrin, G.: South African sign language assistive translation. In: Proceedings of the IASTED International Conference on Telehealth/Assistive Technologies, pp. 7–12 (2008)
168. Zwitserlood, I. et al.: Synthetic signing for the deaf: esign. In: Proceedings of the conference and workshop on assistive technologies for vision and hearing impairment (2004)
169. Blender Tool. https://www.blender.org/features/animation/. Last accessed 2020/09/04
170. Curious Labs POSER. https://curious-labs-poser.software.informer.com/6.0/. Last accessed 2020/09/22
171. ELAN annotation tool. https://www.mpi.nl/corpus/html/elan/. Last accessed 2020/08/09
172. Google Neural Machine Translation. https://en.wikipedia.org/wiki/Google_Neural_Machine_Translation. Last accessed 2020/09/18
173. Humanoid Animation. https://www.web3d.org/working-groups/humanoid-animation-hanim. Last accessed 2020/09/23
174. Maya Tool. https://www.autodesk.in/products/maya/overview?plc=MAYA&term=1-YEAR&support=ADVANCED&quantity=1. Last accessed 2020/08/08
175. Semantic Role Lists. http://elies.rediris.es/elies11/cap5111.htm. Last accessed 2021/04/05
176. Unity 3D. https://unity.com/. Last accessed 2020/08/08
177. VComD:Sign Smith Studio. http://www.vcom3d.com/signsmith.php. Last accessed 2010/08/08
178. WebGL. https://www.khronos.org/webgl/. Last accessed 2021/04/12