IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 24, 2021, accepted June 16, 2021, date of publication June 23, 2021, date of current version July 2, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3091433

A Crowdsourcing-Based Framework for the Development and Validation of Machine Readable Parallel Corpus for Sign Languages

UZMA FAROOQ[1,2], MOHD SHAFRY MOHD RAHIM[1], NABEEL SABIR KHAN[2], SAIM RASHEED[3], AND ADNAN ABID[2], (Member, IEEE)
[1]Department of Computer Science, Universiti Teknologi Malaysia, Johor Baharu 81310, Malaysia
[2]School of Systems and Technology, University of Management and Technology, Lahore 54770, Pakistan
[3]Faculty of Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

Corresponding author: Adnan Abid (adnan.abid@umt.edu.pk)

ABSTRACT Sign languages are used by the deaf and mute community of the world. These are gesture based languages where the subjects use hands and facial expressions to perform different gestures. There are hundreds of different sign languages in the world. Furthermore, like natural languages, there exist different dialects for many sign languages. In order to facilitate the deaf community several different repositories of video gestures are available for many sign languages of the world. These video based repositories do not support the development of an automated language translation systems. This research aims to investigate the idea of engaging the deaf community for the development and validation of a parallel corpus for a sign language and its dialects. As a principal contribution, this research presents a framework for building a parallel corpus for sign languages by harnessing the powers of crowdsourcing with editorial manager, thus it engages a diversified set of stakeholders for building and validating a repository in a quality controlled manner. It further presents processes to develop a word-level parallel corpus for different dialects of a sign language; and a process to develop sentence-level translation corpus comprising of source and translated sentences. The proposed framework has been successfully implemented and involved different stakeholders to build corpus. As a result, a word-level parallel corpus comprising of the gestures of almost 700 words of Pakistan Sign Language (PSL) has been developed. While, a sentence-level translation corpus comprising of more than 8000 sentences for different tenses has also been developed for PSL. This sentence-level corpus can be used in developing and evaluating machine translation models for natural to sign language translation and vice-versa. While the machine-readable word level parallel corpus will help in generating avatar based videos for the translated sentences in different dialects of a sign language.

INDEX TERMS Crowdsourcing, HamNoSys, parallel corpus, sign language dictionary, sign writing.

I. INTRODUCTION

Sign languages are gesture-based languages that are used by the deaf community of the world. There is no universal sign language, and there exist hundreds of sign languages in the word, i.e. every sign language has a different gesture for the same word of natural language [43], [44]. Furthermore, like different written or scripting languages there are different dialects of sign language gestures as well, i.e., in large countries there exist different gestures for the same word in different regions of a country [43], [49]. Figure 1(a) shows different gestures of the same word for Pakistan, British, and American sign languages. While Figure 1(b) presents an example for Pakistan Sign Language (PSL) where it shows the gesture for the word *Lie* for different dialects of PSL.

Each word of spoken language is either represented by a single gesture or a combination of multiple gestures in sign language. The gesture may be static or dynamic which may involve certain movements of hands to perform a gesture. The static signs do not have any movement, thus picture-based repositories are suitable for static gestures. However, the dynamic gestures are not easy to represent and understand in pictorial form. For example, the picture-based representation of static gesture for the word *father* is

The associate editor coordinating the review of this manuscript and approving it for publication was Ali Shariq Imran.

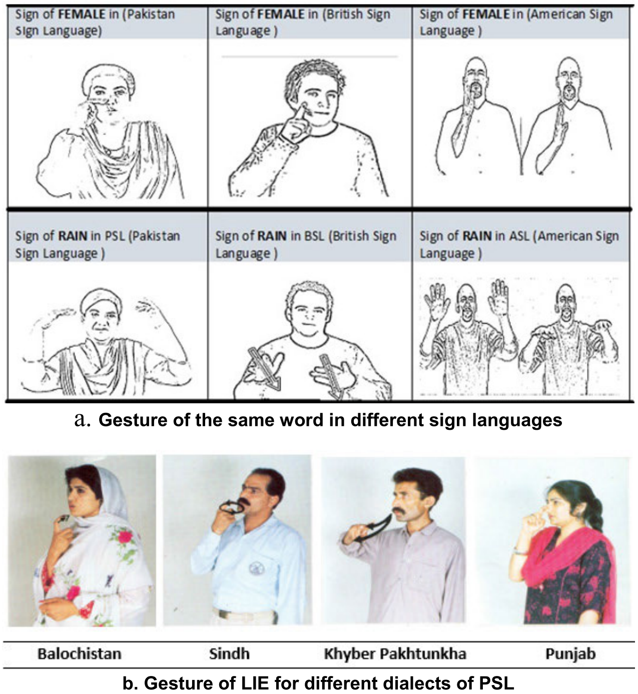


FIGURE 1. **(a) Differences in sign language; (b) dialects of a sign language.**

shown in Figure 2(a). The example of dynamic gesture for the word *car* is shown in Figure 2(b), these gestures belong to PSL. From the figure, it is quite clear that a single word may contain multiple gestures along with the certain type of movements to convey the actual meaning of the sign. Multiple parts of a given sign are represented in separate images and movements are represented with the help of arrows.

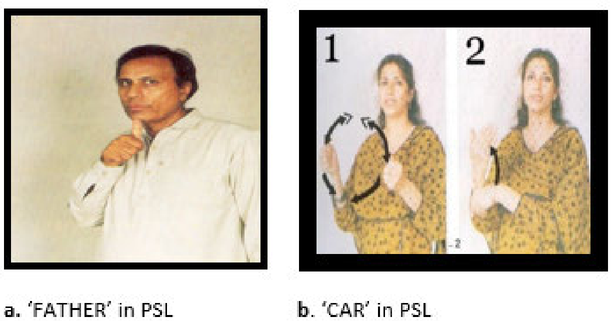


FIGURE 2. **Image-based dictionary: Signs of FATHER and CAR in PSL.**

A variety of tools and applications have been developed for different sign languages by the researchers to facilitate the deaf community of the world in learning a sign language and for building multi-disciplinary sign language translation applications [43], [49], [50]. Among the sign languages, American Sign Language (ASL) [1] is the leading sign language, not only concerning the linguistic details but there exist large multi-media corpora and dictionaries along with supportive tools and technologies. Apart from this, a lot of research work has been accomplished on different sign languages in other regions of the world as well. For instance, British Sign Language (BSL) [2], German Sign Language (DGS) [3], and Spanish Sign Language (LSE) [4] are among the well-studied sign languages in Europe. While recently significant development has been observed for sign languages of many other regions of the world as well including South African Sign Language [5], Indian [6], Vietnamese sign language [7], Bangladeshi [8], Pakistani [9], [13], [14] Thai [10], Arabic [11], and Malaysian [12] sign languages.

A. MOTIVATIONS

As far as the evolution of sign language dictionaries and repositories for gestures is concerned, initially, some image or picture-based repositories were published in the form of books. But they are not easy to understand and are unable to offer automated search and are not easily extendable. Subsequently, the image-based repositories were replaced by video-based corpora. Video-based corpora, as the name suggests, operate by storing a video against each word. These videos are recorded with the help of human signers. Video-based representations for the words *car* and *mother* in PSL are shown in Figures 3(a) and 3(b).

FIGURE 3. **Video-based dictionary of signs Car and Mother in PSL.**

The video representation of gestures enhances the visualization and understanding of both static and dynamic signs as compared to images because movements can be easily seen as a continuous video stream, thus enhancing the understandability of gestures. Currently, there exist many video gestures-based dictionaries for many sign languages of the world. However, there are some drawbacks of video-based repositories as well, as they consume large storage space, and are hard to extend, as they require specialized settings for recording videos. While the biggest drawback of image and video-based repositories are that they are not suitable for developing sign language translation systems, because such systems need to render the gestures for different gestures in a specific sequence. These images/videos may or may not be performed by a single person i.e., one word by one signer, and another word by yet another signer, those too in different settings. Where, the settings include clothes, gender and the environment where videos were recorded. Figure 4 shows the translation of the PSL sentence ''Mother car drive''. The problem is quite clear that these gestures were performed by two different signers even of different genders. Also, the clothes of the same signer vary in the same sentence. This

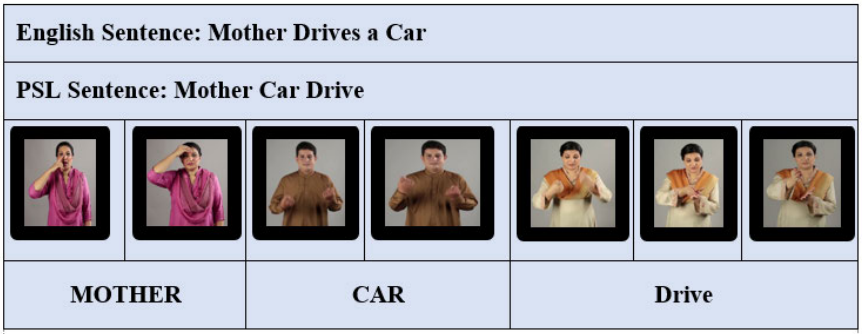


FIGURE 4. Video-based sentence translation in PSL.

TABLE 1. Pros and Cons for video and image-based storage formats.

Storage Type	Pros	Cons
Video Signs	Realistic Easy to create	Time-consuming to create High memory consumption Not suitable for the sentence translation system Not easily extendable
Pictures/Image	Very little memory consumption	Time-consuming to create Not realistic as compared to videos Not appropriate for sentence-level translation system Not easily extendable

TABLE 2. Differences between English and PSL sentence structures.

English Sentence	Equivalent PSL Sentence
Mark is writing a letter.	Mark letter write now
Do you need a book?	Book need you Yes-No?
I am not hungry.	I hungry not.

will seriously affect the usability of the system. A comparison of video and image-based sign language dictionaries has been presented in Table 1.

There exist a large number of sign language dictionaries for different sign languages which contain the gestures for different words and phrases. Some of the advanced countries have developed rich dictionaries for their sign languages, however, in many countries the dictionaries comprise of few hundred gestures for the commonly used words. Thus, there is a need to develop a mechanism that can help to build a sign language dictionary for every country. Furthermore, these dictionaries need to cater to a natural requirement of maintaining dialects for different regions. None of the dictionaries stores regional dialects of words.

The above discussion apprises the need to develop more enriched repositories of sign language that are not only suitable for learning and practicing a sign language but can also help in translation systems by storing the gestures in machine-readable formats [15]. Similarly, there is a need to devise a well-defined process to develop, maintain, standardize, and extend sign language dictionaries and repositories preferably while involving the deaf community and language experts by providing them an appropriate crowdsourcing platform, as recently highlighted by some active researchers in this area [43], [49].

Apart from the word and gesture level details of a sign language, it is interesting to note that the sentence structure of the sign languages is also different from written and spoken natural languages. Table 2 presents some sample sentences in English with their equivalent translated sentences into PSL. It can be observed that after translation the word order has changed, some words have been eliminated, and some additional words have been added in certain cases. This makes the natural to sign language translation an interesting and different machine translation problem as compared to conventional language translation problems. Interestingly, the machine translation problems require a substantial set of training and testing data for building a suitable translation model and testing it properly. Thus, apart from building a word-level machine-readable corpus, there is a need to build a sentence-level repository that comprises a variety of source and target language sentences to develop and test sign language translation models. Furthermore, incorporating multiple languages in the repository can help to develop multilingual translation systems for sign languages.

Another motivation for engaging the crowd for building such repositories is to collect a variety of signs for the same word to build a data-set for the development and testing of gesture recognition systems with the help of multiple signers. It can also help in gathering gestures for different dialects of a sign language.

Similarly, such platforms can also provide a foundation to the standardization authorities to engage the deaf community to define gestures for new words and terminologies, as well as, standardize the existing gestures by considering the most frequent gestures for a word.

Lastly, the deaf community can be engaged to validate the translation systems developed for the deaf community with the help of avatars for the correctness of translation and acceptability of avatars.

B. CONTRIBUTIONS

The major contributions of this research are as follows:

The article presents the following contributions:

A. A crowdsourcing based framework for building a word level parallel corpus, and sentence level translation corpus for sign languages

B. It further presents a process to develop a word level parallel corpus for different dialects of a sign language; and a process to develop sentence level translation corpus comprising of source and translated sentences.

C. The proposed framework has been successfully implemented and engaged different stakeholders to build word and sentence level corpus.

D. Developed a corpus for Pakistan Sign Language using the proposed framework

a. As a result, a word level machine-readable parallel corpus comprising of the gestures of almost 600 words of Pakistan Sign Language (PSL) has been developed.

b. While, a sentence level translation corpus comprising of more than 8000 sentences for different tenses has also been developed using the process outlined in the proposed framework.

The rest of the article has been structured in the following manner: the existing research work has been discussed in Section II. While Section III presents the proposed framework for crowdsourcing-based data collection for building word-level sign language corpus and sentence level translation corpus. The characteristics of the corpus generated from the process have been presented in Section IV. The effectiveness of the word level and sentence level corpus in developing and testing a fully automated sign language translation system has been discussed in Section V. Discussion about the importance of the proposed framework and generated corpus has been presented in Section VI. While, Section VII concludes the article.

II. RELATED WORK

This study presents some details of dictionaries and repositories made and exposed by different sign language experts and organizations across the globe. A British Sign language corpus consisting of 2,528 video clips was proposed in (2011) by [16]. These videos were recorded by the Deaf people using BSL. The corpus is also exposed to the public so that they can use and learn BSL. The European Cultural Heritage Online organization (ECHO) published corpora consisting of children's stories and poetry in Swedish, British Netherlands SL. This corpus contains video signs performed by a single signer. Similarly, [17] proposed a corpus for deaf children in Africa.

Various multilingual dictionaries are having a different kinds of representations are available for different SL. Spanish Sign Language- Spanish (DILSE) dictionary was proposed by [18]. It is a multilingual dictionary is available online for the deaf community of Spain. The dictionary provides two levels of search in which users can either search a Spanish word against a sign and similarly can do a signed search by giving a Spanish word as input. The Italian sign language dictionary was proposed by [19]. An electronic multimedia dictionary consisting of signs of 3 different sign languages including American Sign Language (ASL), Japanese Sign Language (JSL) and Korean Sign Language (KSL) was created by [20]. A Danish Sign Language (DTS) dictionary was developed by [21]. The dictionary stores words along with their synonyms and corresponding human recorded videos of each sign.

Some groups worked on domain-specific SL corpora, details of some of the domain-specific SL dictionaries are described here. One domain-specific dictionary includes a weather reports corpus comprising of 2468 sentences in German and DGS and has been reported by [22]. An ISL Dictionary for disaster domain was proposed by [6]. The videos are traced to convert them into avatar animation. This dictionary provides the information of disaster to the deaf. It consists of 600 sentences and 2000 words. Some dictionaries are enriched with sign writing notations, which help them in showing the gestures using avatar technology. For instance, a DGS Corpus was developed by [23] using gloss and Hamburg Notation System (HamNoSys). Apart from this, very few repositories target machine translation for sign languages. As an example, in [24] the researchers present statistical machine translation experiments on a corpus of about 2000 sentences for the language pair Chinese and French.

There are many sign language dictionaries available online. Repositories according to different regions are along with a total number of videos, format, URL and sign writing notations have been listed in Table 3.

It can be observed that most of the dictionaries have video representations of signs but some contain an image as well. These dictionaries contain signs of words and some of them e.g. ''life print'' also include signs of common phrases. There are also few YouTube channels like ''Elma Production'' that are uploading sign language lessons.

A. SIGN LANGUAGE DICTIONARIES

Experts and communities from different countries have developed sign language repositories that largely comprise sign

TABLE 3. List of sign language dictionaries.

Language	Total word	Format	URL	Sign Writing Notation
American Sign Language (ASL)	Over 2650 Signs	Image/ Video	https://www.lifeprint.com/	
	Over 7835 Signs	Video	https://www.handspeak.com/	graphemes
	10 lessons containing multiple signs	Video	http://www.signlanguage101.com/	
	Over 9273 Signs	Video	https://www.signingsavvy.com	ASL Gloss
British Sign Language(BSL)	Over 21,000 Signs	Video	https://www.signbsl.com/	
	Over 2500 signs	Video	https://bslsignbank.ucl.ac.uk/	
	Over 478 Signs	Image	http://www.british-sign.co.uk	
German Sign Language(DGS)	Over 6000 Signs	Video	www.dgs-korpus.de (in development phase)	
	Over 15000 Signs		https://www.spreadthesign.com	
Spanish Sign Language (LSE)	9 lessons containing multiple signs		http://www.cervantesvirtual.com/seccion/signos/	
	Over 15000 Signs		https://www.spreadthesign.com	
Indian Sign Language(ISL)	Over 2775 Signs	Video	http://www.indiansignlanguage.org	
	Over 3000 Signs	Video	https://www.youtube.com/channel/UC3AcGIlqVI4nJWCwHgHFXtg	
	Over 898 Signs		https://www.talkinghands.co.in	
Pakistan Sign Language(PSL)	Over 7000 Signs	Video	http://www.psl.org.pk	
Arabic Sign Language(ArSL)	Over 2861 Signs	Video	http://sldictionary.appspot.com	
Chinese Sign Language(CSL)	Over 542 Signs	Video	https://www.newsigns.jp/fsle/china	
	Over 44 Signs	Video	http://www.deafstudies.jp	
Malaysian Sign Language (MSL)	13 lessons containing multiple Signs	Video	https://www.youtube.com/user/ElmaProduction/feed	
		Video	https://books.google.com.pk/books/about/MySLang.html?id=C89NAQAACAAJ&redir esc=y	
Australian Sign Language (AUSLAN)	Over 7983 Signs	Video	http://www.auslan.org.au/	
	Over 345 Signs	Video	https://www.youtube.com/channel/UCjCBX ZL4T0nCyq3AgFVsDg	
South African Sign Language (SASL)	Over 800 Signs	Video	https://www.realsasl.com	
	Over 61 Videos having multiple Signs	Video	https://www.youtube.com/channel/UCpuibTErIqB3oEvp0Jaap9w	
Brazilian Sign Language (LSB)	9,500 entries	Images	http://www.signwriting.org/brazil/brazil21.html	

language dictionaries. The repositories are also referred to as sign language dictionaries, corpus, and at times corpora. Most of these dictionaries comprise of most frequently daily used words in different useful multimedia formats. The most common representation in sign language dictionaries is a presentation of words in the form of pictures, videos, or in the form avatar videos. In terms of understanding the gestures, the video representations are much easier as compared to the pictorial representations. While, the avatar-based video representations are also understandable, yet they are considered less expressive as compared to the videos. However, in terms of data storage, the pictorial representation takes least storage space, while avatar videos take more storage, and the human videos take even more storage space. Some dictionaries have also stored the gestures in sign writing notations, which helps to store these gestures in a textual representation that can be later converted into a gesture using avatar technology. Certainly, the textual format takes the least amount of storage but requires tools to convert it back into the equivalent sign language gesture. More details related to sign writing notations and avatar technology have been discussed in the coming sections.

These dictionaries also vary in terms of granularity of language components. For instance, almost all of them include the gestures for all the letters and basic numbers for the considered sign language; the next granularity is that of storing the gestures for words i.e. they store words of a spoken language along with its representation in the considered sign language. Many dictionaries also store some frequently used basic phrases e.g. greetings etc. which is another level of granularity. Furthermore, many languages have exposed the sign language based children stories and poems which comprise of several sentences, which is yet another higher level of granularity. Most of these dictionaries present the information under different categories, and are augmented with searching features.

B. LIMITATIONS OF SIGN LANGUAGE DICTIONARIES

Following are the limitations of multimedia-based representation of sign language gestures which invite the researchers to devise methods to develop suitable sign writing notations.

1) STORAGE SIZE

The size of storing a gesture in image or video format requires a substantially large amount of storage as compared to text-based storage. These challenges the scalability of large-scale systems.

2) LACK OF REPRESENTATION IN IMAGE-BASED GESTURES

The dynamic gestures are hard to represent through image-based storage, as they are presented through multiple frames showing different hand shapes, orientations, while the movements are shown with the help of different symbols like arrows etc.

3) STANDARDIZATION

There are two different notions of standardization. The first one is that of standardization in the recording of video or image-based gestures. This may involve engaging the same person for performing the gesture while maintaining the same background settings, clothes etc.

While the other notion deals with variations of signs for the same word, due to different dialects or subjective differences. In this case, the same word or phrase can be signed by different people in different ways, or the gesture of the same word may vary from one region to another for the same sign language. To this end, almost every country has a responsible organization that works for maintaining the standard national sign language gestures e.g. Americans National Institute of Deafness and Another Communication Disorder (NIDCD) is responsible for ASL [25]; Pakistan Directorate General of Special Education Initiative National Institute of special education (NISE) [26] are responsible for managing PSL; while Indian National Institute of the Deaf [27] does it for ISL.

4) ISSUES IN SENTENCE GENERATION

The gestures are connected in a specific sequence to reflect a meaningful sentence. This can certainly be done by rendering the gestures of these words in a given order. However, this poses a lot of challenges while using an image or video-based repository. For example, gestures recorded by different persons in different settings, or by the same person with different clothes result in the user experience of the translation system. While in the case of an image-based repository it becomes even more tedious where different frames are loaded while showing the arrow signs representing the movements. Figure 4 shows the generation of different sentences of PSL using video gesture repositories. It is clear that gestures for the words involve in a sentence have been performed by different persons, thus making it difficult to understand.

Another important aspect of sign language dictionaries is the standardization of gestures for each language. Since every region has its particular sign representation that varies from region to region, and it is pertinent to enrich and extend these dictionaries to include regional dialects of these sign language. Furthermore, it is imperative to have a standard process for the inclusion of a gesture into the national language.

C. USE OF CROWDSOURCING FOR NATURAL LANGUAGE PROCESSING TASKS

Researchers have discussed the use of crowdsourcing for different types of natural language processing tasks [28], [46]. In particular, a high-level approach for building a crowdsourcing-based sign language dictionary has been discussed in [29] and [30]. Similarly, general guidelines However, there is a need for a well-defined crowdsourcing platform that engages the deaf community, language experts, and technologists to build a multi-purpose sign language dictionary that can be used for building specialized translation tools for the deaf community. Similarly, different successful usages of crowdsourcing in relevant projects like the Bentham project [37], workforce-efficient consensus for bio-collections information [38], use of crowdsourcing in general [39], and application of crowdsourcing in corpus management in natural language processing [40], [45] has also been presented in the literature. On the other hand, researchers have highlighted the quality assurance issues in general-purpose crowdsourcing platforms like Amazon Mechanical Turk [47], [48] which questions its suitability for NLP tasks. Similarly, special purpose and language-specific crowdsourcing platforms have also been proposed e.g. [46] proposed a crowdsourcing platform for collecting and annotating Arabic language tweets.

In particular, to sign languages, an idea with an elementary implementation of a web-based tool for building a parallel corpus was proposed by Becker *et al.* [42]. This tool involves users annotating a piece of text or a document with the help of users who fetch already stored gestures for a word. But this tool used SignWriting notation for storing the gestures, thus an annotator requires the understanding of this notation for annotation. Furthermore, SignWriting cannot be used to generate automatic gestures for sign language. Likewise, there is no editorial check on the quality of the output generated through this tool. Lastly, it does not help to differentiate the sentence-level grammatical differences and variations between sign language and natural language sentences.

The literature review clearly shows that sign language generation, recognition, and translation involve multidisciplinary perspectives of linguistics, gesture recognition, avatars, and translation system. Whereby, the translation system requires corpora comprising of natural language text to sign language

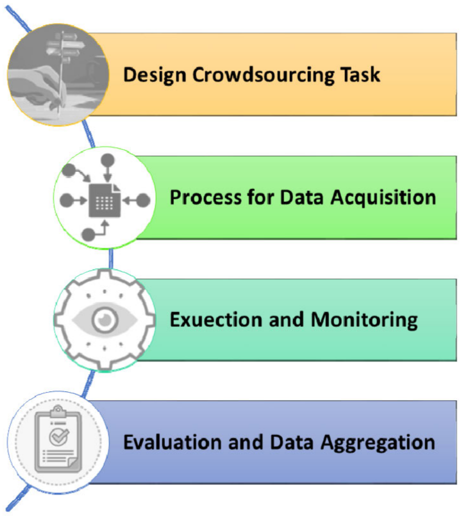


FIGURE 5. Stages of the process for crowdsourcing based NLP tasks.

translations, while gesture recognition requires a huge corpus of images and videos. Thus there is a strong need for a framework that facilitates the creation of these large-scale corpora. To the best of our knowledge, there is no detailed framework that involves crowdsourcing and editorial control for sign language corpus generation and verification for a variety of dialects of a sign language. Thus, this research proposes a generic framework that involves the deaf subjects as contributors and validators, and the sign language experts as editors and reviewers of the submitted work to build a word and sentences level parallel machine-readable corpus for sign languages.

III. A FRAMEWORK FOR THE DEVELOPMENT OF SIGN LANGUAGE CORPUS USING CROWDSOURCING

This section presents the theoretical framework for the development of crowdsourcing based parallel corpus for sign languages. The proposed framework aims to create extendable and standardized machine-readable sign language repository, whereby it employs crowdsourcing for involving the deaf community and language experts as contributors and validators for different tasks. The proposed framework is based on some recommended stages for natural language-based crowdsourcing tasks that involve (*i*) design crowdsourcing task; *(ii)* process for data acquisition; *(iii)* execution and monitoring; *(iv)* evaluation and data aggregation [28], as shown in Figure 5.

Overall, the proposed framework has been divided into two main parts referred to as data submission, and data validation components, as shown in Figure 6. The data submission component is responsible for data acquisition with the help of the crowd. Whereas, the data validation component is responsible for ensuring that only correct data is stored in the system. The framework involves two different granularities of data, firstly it acquires and validates the gestures of alphabets, words, and phrases of a sign language. Secondly, it engages deaf subjects

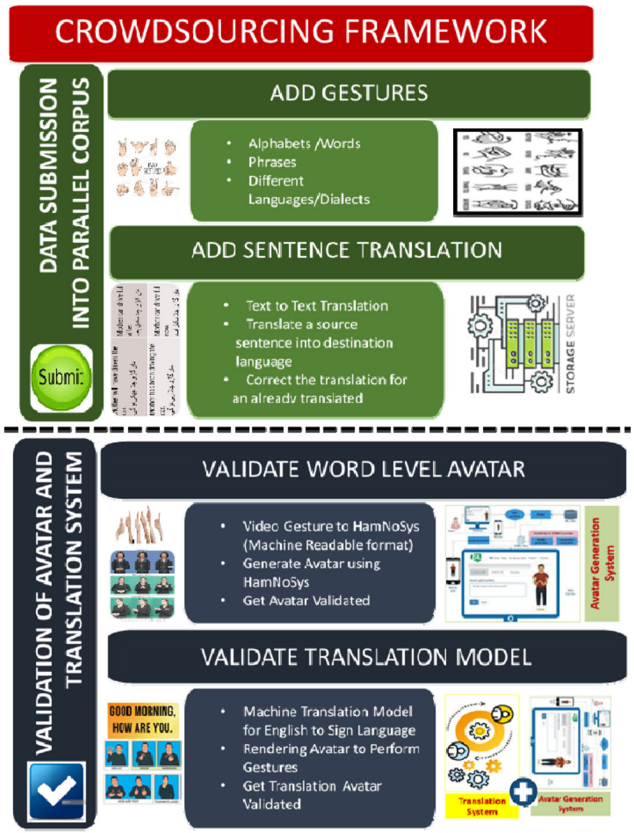


FIGURE 6. Proposed crowdsourcing framework.

to add and validate sentence-level English to sign language translated data.

A. PROCESS FOR DATA ACQUISITION AND VALIDATION

At this granularity level, the proposed framework involves a six-step process as shown in Figure 7. Whereby, the first 3 steps help to add a video gesture into the sign language dictionary through an editorial process. While the next 3 steps are responsible for converting the input into a rich machine-readable sign writing notation that is subsequently used for generating an equivalent avatar. A brief description of each step of the process is as follows:

i. Submission of a video gesture for a word in different dialects of PSL by involving deaf subjects as a crowd.
ii. Get the collected gestures reviewed by involving PSL experts as reviewers and editors using a well-defined editorial process.
iii. Add correct gestures into the repository.
iv. Convert gesture to a rich machine-readable format.
v. Convert machine-readable gestures into a video avatar.
vi. Validate avatar by involving deaf subjects.

1) SUBMISSION OF A GESTURE

The process starts with a selection of the most frequent English language words used in daily life communication. The words are grouped in different categories and are exposed to the deaf subjects as a call for gestures. The next step in the process is to register the deaf subjects and sign language

FIGURE 7. Process to add and validate a sign language gesture.

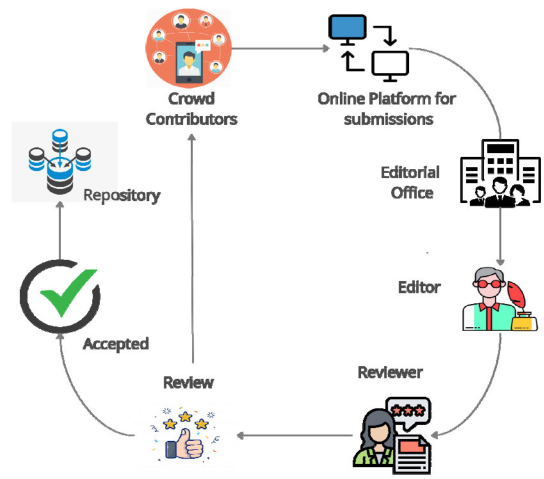


FIGURE 8. Editorial assessment process.

experts from various deaf institutes. The deaf subjects submit the gestures for different words listed in the call for gestures. In this research, 40 students from a deaf school, and a special education department of a university were registered as deaf subjects.

The data acquisition component invites the contributors to submit a gesture by writing or selecting a word or phrase along with the dialect of the language, whereby the contributor can upload the recorded gesture. The selection of dialect and region can help in collecting gestures for the same word in different regions. This will help to store different variants of the gestures for the same word or phrase in a given sign language. Subsequently, the detailed analysis and processing of the uploaded videos and relevant meta-information can help in the standardization of gestures in a language. For instance, after collecting a significant amount of gestures for a given word, we may choose the most widely used representation of a gesture as a standard for that word.

The gesture submission sub-component also provides an option to invite the registered users to submit the gestures for a specific set of words. For instance, to prepare a dictionary for all the words and phrases being used in the English book for Class 1, a call for gestures can be initiated. This can not only help in collecting relevant gestures but can also help to build supportive tools using the collected vocabulary of sign language. Similarly, calls for certain domain-specific gestures can also be initiated. For instance, in low resource sign languages, there might be a need of adding gestures for the vocabulary used in hospitals, shopping malls, train and bus stations etc. Therefore, this component allows initiating a call for gestures for a specific context, where it provides words and phrases for that domain and invites the deaf subjects to contribute for gestures in their regional dialects.

Apart from this, there is also a need for generating new gestures in a standardized manner. Particularly, new terminologies like COVID-19 or any other new technical term require appropriate gestures which are not present in a given sign language. The proposed framework also provides a feature that involves the language experts to create and suggest gestures for new words or terms. The suggested gestures are then discussed and evaluated by all the involved experts to finalize one gesture from all the submitted proposals. This feature does not involve the general public as the new gestures should be developed by the experts of the language.

2) EDITORIAL PROCESS

Most of the crowdsourcing systems lack quality control. Therefore, the proposed framework involves a well-defined review process to review the submitted gestures by involving different sign language experts as editors and reviewers. The whole editorial process has been presented in Figure 8. The whole activity starts with the submission of a gesture by a deaf person who belongs to the contributing crowd. Which, after a preliminary check by the editorial staff is assigned to an *editor*, who in turn, assigns this gesture to one or more language and dialect experts, who are given the role of a *reviewer* in the proposed system. The expert reviews the gesture by reviewing its manual and non-manual features. A reviewer may either accept the gesture, ask the contributor to improve it, or reject it for further consideration. Thereafter, the editor will make the final decision about the acceptance or rejection of gestures. If the final decision is that of acceptance, then the gesture will be added to the repository. Whereas, in case of asking for a revision, the contributor will submit a new video after improving the gesture based on the comments received from the reviewers. Subsequently, it will be marked to the same reviewers for evaluation and can either be accepted or rejected at this stage. Lastly, the rejected gestures will be discarded and shall not be added to the gesture repository.

Figure 9 presents the rubric to facilitate the reviewer for the evaluation of a gesture. It can be observed that the rubric for the evaluation of a gesture has been carefully designed while considering the manual and non-manual features of a sign language gesture. Where the manual features are

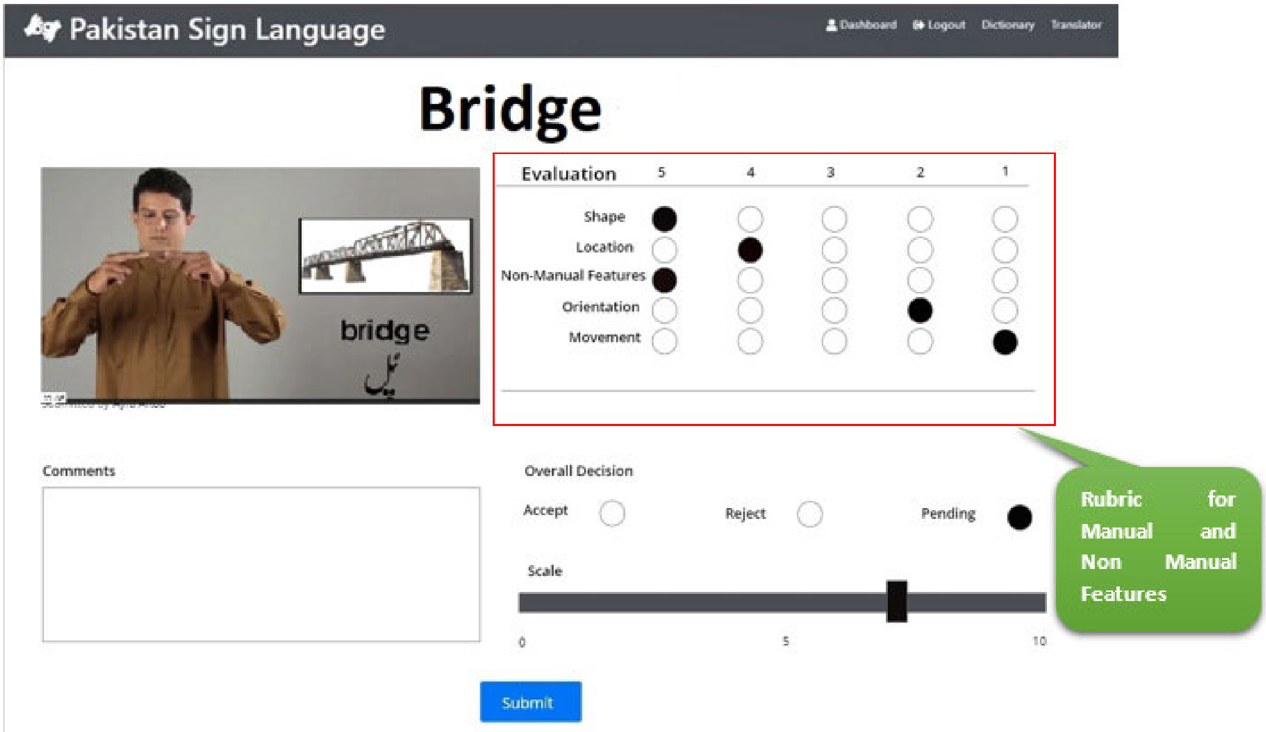


FIGURE 9. Rubric for evaluation of a submitted gesture.

related to hand-shapes, location, orientation, and movements. While, the non-manual features are related to the other body movements including shoulder shrugging, facial expression, mouthing, eye-brows raising etc.

Although some gestures might be simpler and may not involve any movement or non-manual features, yet the provided rubric is based on maximum details. Apart from this, the reviewer can also provide relevant details in a descriptive form in the available text area. Lastly, every reviewer has to rate the gesture on a scale from 0 to 10. This will help in computing an average score for all accepted gestures. Thus, the gesture with the highest score will be shown as a main gesture for a given word, as we may have several gestures for a given word, which may be submitted by many different contributors. Similarly, the selected best gesture is used to convert the gesture into equivalent sign writing notation.

3) SAVE GESTURES

In this research gestures for 500 different words have been stored. The analysis of video data collected revealed that there were 1328 videos recorded by a team of 50 students. Further analysis depicted that about 389 words have 2 videos each, 73 words have 3 videos and 35 words have 4 videos each.

4) CONVERSION OF GESTURE INTO MACHINE READABLE FORMAT

The notion of parallel corpus involves multiple useful representations of the words in the dictionary or corpus. The proposed framework initiates the data collection by writing or selecting a word or a phrase. There are many different representations of a word e.g., English version, Roman Urdu version, Urdu language version, its video gesture uploaded by the contributor. While after accepting the gesture, the technical team is responsible for converting the video gesture into an appropriate machine-readable sign writing notation, which can subsequently be converted into an avatar. We may also store the avatar-based representation. Therefore, it is pertinent to store the machine-readable form of each gesture as well.

Like spoken languages Sign languages are also be transcribed using various sign writing notations [15]. There are many notation systems used for Sign language writing among which the four most widely used Sign Writing Notation Systems are Stokoe, Gloss, Sign Writing, and HamNoSys [31], [32]. The basic representations of widely used sign writing notation symbols are shown in Figure 10. While a comparative analysis of these widely used sign writing notations has been presented in Table 4.

After the selection of HamNoSys as the sign writing notation now the challenge is how to generate HamNoSys for our repository. For this purpose, a dedicated team of students work using a well-defined process, to generate and validate HamNoSys, as shown in Figure 11. It is imperative to discuss that the gesture repository may have more than one acceptable video for a given word. Thus, the first step would be to select the best-rated video gesture for every word so that it can be used to write the HamNoSys.

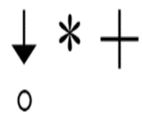
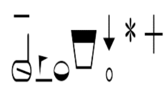
5) GENERATION OF AVATAR

The first step of the HamNoSys writing process requires careful identification of hand shapes, location, movements,

TABLE 4. Comparison of Widely used Sign Writing Notations.

Sign Writing Notation System	Sign Language Dependent	Non-Manual Features Support	Objective	Arrangement	Computer Compatibility
Stokoe	Yes	No	Dictionary or Academic	Linear	Custom Font or ASCII codes
GLOSS	Yes	Yes	Academic	Linear	Custom Font or ASCII codes
SignWriting	No	Some	Public Use	Pictorial	ASCII or Unicode
HamNoSys	No	Yes	Academic	Linear	Custom Font Unicode

TABLE 5. Mapping of hand shape, location, orientation, and movement using HamNoSys.

Sign in PSL	HamNoSys				
 Today	**Hand Shape**	**Hand Position**	**Hand Location**	**Hand Movement**	**Complete Vector**
					

and non-manual features. These features are then mapped onto the respective HamNoSys symbols using the HamNoSys keyboard. The HamNoSys can be converted into SiGML which is an intermediate mark-up language representation that can further be converted into an avatar that performs the gesture based on the input HamNoSys. The generated avatar is reviewed by the responsible person and after some iterations of refining the avatar, an appropriate avatar is generated. The number of iterations depends upon the complexity of the gesture being transcribed as well as on the expertise of the HamNoSys writer. However, it is an offline step and the concerned person can write correct HamNoSys in a few iterations. Once the avatar is ready, it is evaluated by a sign language expert who may either accept the avatar or may require more refinement in the avatar, which can be done by refining the HamNoSys for that gesture.

The HamNoSys vector is stored corresponding to each word which will later be used to generate the avatar. Table 5 shows the mapping of hand shape, location, orientation, and movements for the gesture of word today in PSL using the HamNoSys.

To figure out the correctness of generated HamNoSys vector it is required to find a mechanism through which some kind of visual output can be generated to see how accurately the sign can be performed by the avatar. In this research, third-party software is used which takes HamNoSys as input and automatically generates an intermediate representation called SiGML. The tags are similar to XML tags which are read and played by a by SIGML player.

6) VALIDATION OF AVATAR

The Avatar output is verified by SL experts again. There are few cases in which the output was not correct due to wrong mappings of HamNoSys symbols. In these cases, the problems were identified and HamNoSys was again generated till the avatar output is verified by SL experts.

B. TRANSLATION CORPUS: SENTENCE LEVEL DATA COLLECTION AND VALIDATION

To develop a sign language translation corpus, it engages the crowd to translate different types of sentences into equivalent sign language sentences.

The process to collect sentence-level data has been presented in Figure 12. It shows that a variety of English language sentences will be selected for translation into equivalent sign language text. These sentences will be assigned to different deaf subjects for translation. Later on, the translated sentences will be collected and stored in the sentence repository.

1) SOURCE SENTENCES

The source sentences can be selected based on different types of English sentences. Figure 13 shows different possible types of English language sentences. These sentences can be categorized based on tenses into the present, past, and future tenses. Whereas, each tense has four different variants indefinite, continuous, perfect, and perfect-continuous. Similarly, concerning the meanings, the sentences can be categorized

FIGURE 10. Different sign writing notations.

FIGURE 11. Process to Generate and Validate a HamNoSys.

into imperative, negative, and interrogative sentences. While, structurally the sentences are categorized into simple, compound, complex, and compound-complex sentences.

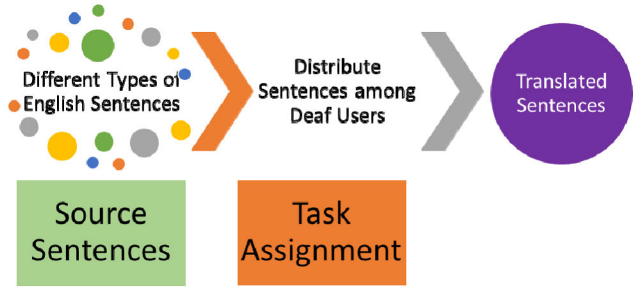


FIGURE 12. Process to develop sentence-level translation corpus.

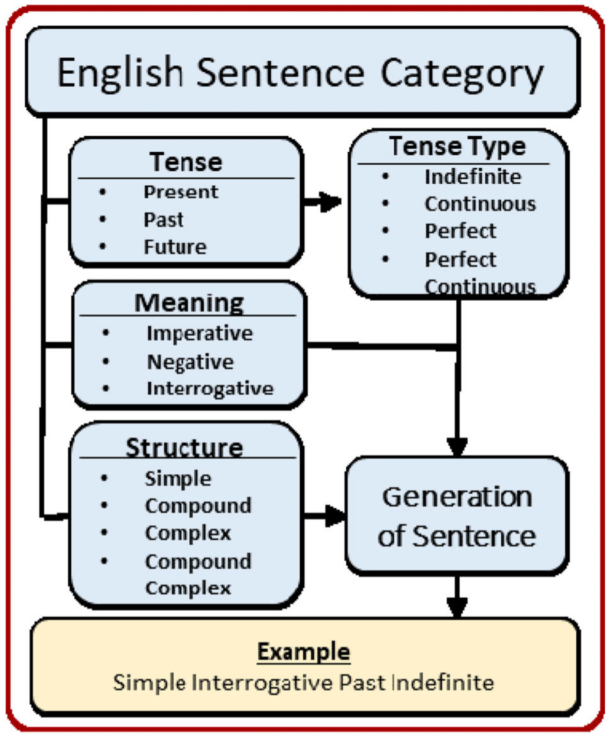


FIGURE 13. Categories of English language used to build sentence-level translation corpus.

In this research more than 8000 source sentences have been collected from different books including the practice books for English grammar and composition, and some textbooks from the primary school syllabus.

2) TASK ASSIGNMENT

The source sentences are assigned to different registered users who are willing to support in generating translation data. These users translate each sentence by themselves. The translated data is collected and stored in the system. This data will be helpful in training and testing different sign language translation systems. There are very few corpora that facilitate in developing and testing machine learning or deep learning-based sign language translation systems. This component of the proposed framework can help to gather translation data for a variety of sign languages by assigning

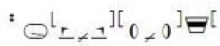
English	Urdu	Roman Urdu	HamNoSys	Human Video	Avatar Video
Car	کار	Kaar			

FIGURE 14. Representation of a gesture in Parallel Corpus.

different source sentences to the users of different sign languages.

In this research, we have collected translations of more than 8000 sentences with the help of users, who translated a variety of English language sentences into equivalent PSL sentences in text format. Some of the sample sentences have been presented in Table 5.

3) VALIDATION THROUGH SENTENCE LEVEL CORPUS

This subsection discusses how the sentence level corpus can be used for measuring the accuracy of any machine translation model for natural to sign language translation. Furthermore, it can also help evaluate the sentence level translation performed with the help of avatar by integrating it with any translation system which uses a word-level machine-readable representation of gestures discussed in the previous section.

The sentence-level corpus is used to evaluate the accuracy of any sign language translation system. This can be helpful for both ways translations i.e. natural to sign language translation, and also for gesture recognition and converting it into natural language sentences. Particularly, for PSL, a natural to sign language translation model has been proposed by Khan *et al.* in [41]. The manually translated output helps us in evaluating the translation system by comparing the translation output with the manually translated sentence available in the corpus while involving measures like BiLingual Evaluation Understudy Score (BLEU Score) and Word Error Rate (WER).

On the other hand, the same translation system helps generate semi-automatic sentence-level translation. Whereby, the involved subjects are shared the English language sentence as well as the translated sign language sentence with the help of the translation system proposed by Khan *et al.* in [41]. In this case, the subjects have two options either to simply mark a sentence as correctly translated, or make the necessary change to the translation. This in turn provides two benefits, firstly it saves time in generating translation corpus; and secondly, it helps to identify errors in the translation system which can be rectified later on.

Lastly, the translated sentences are converted into an avatar with the help of a word-level machine-readable corpus. As the gestures of the translated sentences can be rendered from the gesture-based corpus discussed in the previous section, thereby facilitating the avatar-based translation. Figure 15 shows that the avatar-based video for a PSL sentence looks much better than the video or image-based rendered video, as its look and feel do not change while rendering the gestures. Again, to this end, this research seeks support from the translation system proposed by Khan *et al.* in [41].

TABLE 6. Statistics of word level data in the parallel corpus.

Total Gestures	**~700+**
Letters (English)	**26**
Letters (Urdu)	**37**
Numbers	**25**
Words	**600+**
Phrases	**17**

IV. CHARACTERISTICS OF CROWDSOURCING BASED PARALLEL CORPUS FOR PSL

The first-ever parallel corpus of PSL was proposed in this research. The corpus comprises of two parts: the first part contains gestures and other relevant information for more than 700 different words used in sign language; while the second part comprises more than 8000 sentences translated from English to PSL.

The details related to the persons involved in different roles to build this proof-of-concept application to justify the proposed crowdsourcing framework have been presented in Table 6. Furthermore, the statistics related to the letter/alphabets, numbers, words, and phrases generated by employing the processes defined for the data collection framework have also been presented in Table 6.

A. PARALLEL CORPUS FOR GESTURES

The parallel corpus stores multiple representations of English words including Urdu and roman equivalent text, machine-readable representation of PSL gestures of a corresponding word using HamNoSys, video recorded by human signer along with the iconic symbol of the word. Lastly, the corpus stores the avatar-generated video from the generated HamNoSys vector and SiGML. Figure 14 shows an entry for the word *car* in a parallel PSL corpus.

TABLE 7. Sample sentences for generating translation corpus.

Category of Sentence	English Sentence and Structure	PSL Sentence and Structure
Present Indefinite (Tense)	Peter goes to the market	Peter market go
Present Continuous (Tense)	Peter is going to the cinema.	Peter cinema go now.
Present Perfect (Tense)	Peter has gone to the shop.	Peter shop go full.
Future Indefinite (Tense)	Peter will not read the book.	Peter book read not after
Past Indefinite (Tense)	This was a good match.	This match good was.
WH-Interrogative (Meaning)	Why is Peter going?	Peter go now why?
Aux Interrogative (Meaning)	Do you need any help?	You help want Yes-No?
Compound (Structure)	Car is good and the ride is comfortable.	Car good and ride comfortable.

The corpus consists of various categories of words. Initially, about 500 words taken from basic grammar books and the ones used in daily life belonging to different categories have been selected for data collection. The categories include animals, body parts, week days, color names, relationships and few others. Apart from this, gestures for letters of English and Urdu alphabets, and basic numbers were also collected, which collectively sums up to more than 700 gestures, as shown in Table 6.

The data of these gestures are placed in their respective categories. This gathered pool of English words are categorized in different basic domains like relationship names, school and college most often used words, color names, month names, etc. New categories along with words can be easily added using the corpus insertion interface.

A team of 55 persons was involved to materialize this idea of crowdsourcing-based corpus generation. This includes 2 Editors, 8 Reviewers, 40 deaf students including 32 males and 8 females, while 5 technical resources were engaged to convert the collected gestures into HamNoSys.

B. TRANSLATION CORPUS

The translation corpus has been developed by involving the crowd whereby they were provided with English sentences and were asked to translate them into equivalent PSL sentences. Table 7 shows some sample sentences translated by the involved subjects.

Whereas, Table 8, IX, and X show different useful statistics of the sentence level translation corpus for PSL using the crowdsourcing framework.

It is pertinent to mention that the gestures of the translated sentences can be rendered from the gesture-based corpus discussed in the previous section, thereby facilitating the avatar-based translation. It is pertinent to mention that most of our deaf subjects are studying in university and were comfortable in translating sentences from English to PSL. To the best of our knowledge, this is the first-ever machine-readable corpus for PSL which can be used not only for word-based translation but also includes sentence-level translation.

TABLE 8. Statistics of sentences used to build translation corpus.

Tense	Number of Sentences
Simple Present	500
Present Indefinite	650
Present Continuous	650
Present Perfect	650
Present Perfect Continuous	650
Past Indefinite	650
Past Continuous	650
Past Perfect	650
Past Perfect Continuous	650
Future Indefinite	650
Future Continuous	650
Future Perfect	650
Future Perfect Continuous	650

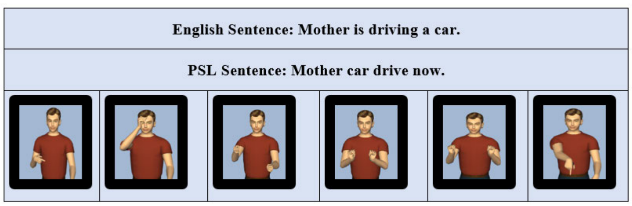


FIGURE 15. Video of translation using Avatar.

Figure 15 shows the translation of the PSL sentence Mother Car drives now using the parallel corpus. It can be observed that the avatar-based video for a PSL sentence looks much better than the video or image-based rendered video, as its look and feel do not change while rendering the gestures.

TABLE 9. Number of sentences (based on meaning).

Meaning	Number of Sentences
Imperative	4500
Interrogative	1600
Negative	2200

TABLE 10. Useful statistics of translation corpus.

	English	PSL
Maximum Length of a Sentence	16	11
Average Length of a Sentence	5.4	4.2
Distinct words (vocabulary) in all sentences	987	634
Total Number of Sentences	8300	8300

To this end, this research seeks support from the translation system proposed by Khan *et al.* in [41].

V. EVALUATION FOR EFFECTIVENESS OF CROWDSOURCING BASED PARALLEL CORPUS FOR SIGN LANGUAGE TRANSLATION

This section discusses the effectiveness of using a machine-readable corpus for sign language translation systems. The principal motivation behind the development of the parallel corpus is to use it for language learning and translation purpose by capitalizing on its ability to be machine-readable. Thus, the evaluation for the effectiveness is primarily based on the comprehensibility and usability of the translation systems that can be developed using avatar technology with the help of the corpus. Whereby, the comprehensibility refers to the richness of the avatar for the sake of understanding it; while the usability aims to gauge the general applicability of the avatar-based translation system by rating it on a scale of 10.

Apart from this, another anticipated advantage of such a machine-readable corpus over the image and video-based is low memory space consumption in storing the representation of signs. Furthermore, it also offers a significantly less processing time for the dynamic generation of sentence-based translation.

A. EVALUATION OF THE EFFECTIVENESS OF THE AVATAR

The generated avatar is performing PSL gestures for the words in their given sequence. This inevitably invites us to evaluate the effectiveness of the avatar in terms of its *comprehensibility* and *usability*. Comprehensibility gauges the richness of the avatar to make the user easily understand the performed gestures [31], [36]. It involves naturalness, rhythm, expressiveness, scale, position, and contrast. While, *usability* means the general applicability of the avatar-based system, where the users were supposed to rate the system on a scale from 0 to 10, where 10 means a perfect system.

1) EVALUATION OF THE COMPREHENSIBILITY

Ten subjects including deaf students and interpreters were involved to test the comprehensibility of the generated avatar. All of them were involved in the data acquisition and were well aware of the objectives of this research. They were also eager to use this new technology-based solution and were comfortable in using these technologies.

a: NATURALNESS

Almost all of the participants felt uncomfortable with the stiffness of the virtual character. Certainly, the avatar-based character cannot perfectly look like a real person in the video. The major concerns were in the movements of fingers and wrist while performing different hand orientations and shapes. Nevertheless, they assured that the avatar is understandable, though it can be improved further.

b: RHYTHM

The speed of performing a gesture by the avatar was also a problem noticed by many people involved in the experiment. The speed of performing the gesture is controllable and can be adjusted by each user. Therefore, this problem was solved in most cases.

c: EXPRESSIVENESS

The expressiveness mostly involves non-manual gestures including the movement of eye brows, lips, etc. In the current form, our avatar does not support non-manual features. However, they are part of future improvements.

d: CONTRAST AND EMPHASIS

The gestures dominantly involve hand movements, thus the hands of the avatar are significantly different from the color of the clothes. Thus, the hands are visible and help to focus on the manual gesture.

e: POSITION

Initially, the system was showing the full body of the avatar, which was disliked by the deaf community. As the signing space involves the top and middle part of the body. Therefore, this problem was rectified and all the participants showed their satisfaction with the position on the avatar.

2) EVALUATION OF THE USABILITY

The usability of the system was measured by asking the participants to rate the avatar-based translation system for its effectiveness for performing translation on three different levels of granularity: *i)* translation of an alphabet; *ii)* translating a single word; *iii)* translating a phrase. Again the same ten participants were supposed to rate the translation system from 0 to 10 score, 10 being the highest. They were also provided a space to share the strengths and shortcomings at each level

TABLE 11. Average rating scores given by the deaf subjects to letters, words, and phrases.

	Human Video Rendering System	Avatar Based System
Letter /Numbers	9.8	9.3
Word Level (Simple)	9.7	9.2
Word Level (Complex)	9.5	8.9
Phrase Level	9.6	8.9
Sentence Level	8.5	8.4
Paragraph Level	8.5	7.9

of granularity. Where they have presented two outputs for each input word, sentence, or paragraph. First, a system that renders human videos of each word of PSL sentence, while the other automatically generates an avatar to perform translation. The users are supposed to rate each output. We present the comparison of average ratings assigned by the users.

a: TRANSLATION OF ALPHABETS AND NUMBERS

This involved translation of an alphabet or a number into a PSL gesture. It was the simplest activity of just reading a gesture from the repository and play it. The users compared the video gesture with the avatar-based gesture. The users were presented with two random alphabets and numbers. The average rating for word-level translation while rendering human video was 9.8, whereas that of the avatar was 9.3, as shown in Table 11. The general comments reveal that avatars are understandable but need to be enriched further.

b: TRANSLATION OF A WORD

The second experiment was designed to compare the avatar and human videos for different words. It is pertinent to mention that some words have simple gestures i.e. they involve a single-handed gesture with no non-manual gestures. While others involve double hand gestures and some non-manual features as well. Thus we have evaluated both categories differently. The average rating for simple single-handed human video-based gestures was 9.7 while that of avatar-based gestures was 9.2. Whereas, the average rating of a human video gesture of a complex single word gesture (the one that involves both hands and some non-manual features) was 9.5 while that of the avatar-based video was 8.9.

c: TRANSLATION OF PHRASE

The third experiment was to compare the human videos and relevant avatar videos for the commonly used phrases. Once again, in this case, two random phrases were shown to different users and they were supposed to assign a rating to the video-based gesture as well as to the avatar-based gesture. Its results were almost similar to those of complex word gestures. As most of the phrases involve double-handed gestures with some non-manual features. Thus the average score of a human video, in this case, was 9.6 while that of an avatar-based video was 8.9 as shown in Table 11.

d: TRANSLATION OF A SENTENCE

The fourth experiment was to test the usefulness of the avatar-based translation for different sentences. All the users were presented with 5 human and 5 avatar-based videos playing a single sentence. The users were asked to rate their understanding between 1 and 10. The users were more comfortable with the human video as it got an average rating of 8.5, but were excited to see the avatar video as well and awarded an average rating of 8.4 to the sentence level avatars, as shown in Table 11. It can be observed that sentence level translation using video is not rated high by the deaf community. The reason is that videos from different persons in different settings show-up in this scenario. While avatar based translation is very close to the video based translation as it is performing the gestures in a consistent settings.

e: TRANSLATION OF PARAGRAPH

In this experiment, the users were shown the human video (manually annotated) and avatar-based video of a famous childhood story "Thirsty Crow". In this scenario, it was observed that the human video was much more appreciated as compared to the avatar video. The major reason pointed out by the deaf subjects was the missing references of pronouns during the rendering of avatars. The deaf community uses the sign space to register the subjects and objects while performing gestures, and the skilled human signer easily used this feature. However, in the current form, the avatar rendering system only renders the gestures of each word without exploiting this feature. However, this can be incorporated in future research for improving the gesture-performing avatars.

B. PERFORMANCE COMPARISON OF IMAGE, VIDEO, AND AVATAR

The performance analysis based on space and processing time for the generation of PSL sentences using all three approaches was done and results are shown in Figure 16. More than 100 videos, their GIFs and SiGML representations were stored in separate folders. 3 words, 10 words, and 100 words were given as input.

The graph reveals that video loading for 3 words took almost 2 seconds, 10 words took 6 seconds to load, while 100 words took approximately 17 seconds to load from disk to player. GIF based sentence generation performance relatively better than human videos but still for sentence having 100 words it still take 8.8 seconds which is large enough in automated translation applications. However, the processing time for avatar-based rendering through HamNoSys is negligible as compared to the other approaches, as it took only 2.1 seconds to dynamically search and render a PSL text having 100 words. On the other side, videos took around

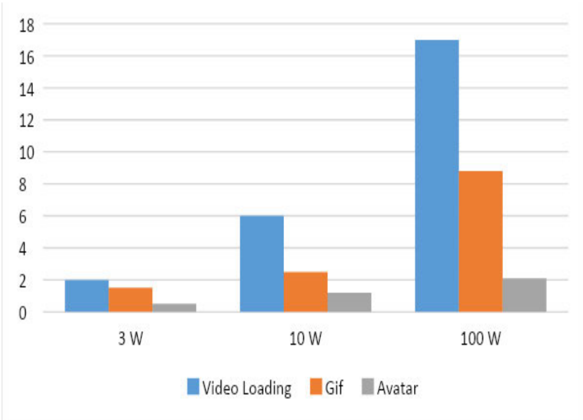


FIGURE 16. Speed of rendering 3, 10, and 100 words using different storage formats.

90 MB, GIFs took 70Mb, whereas the SiGML representations of these 100 words took around 110Kb. From the above experimental results, it is quite clear that our corpus building approach is better both in terms of space utilization and as well as computational efficiency for rendering and playing the avatar for all the words. The corpus is also easily scalable to store more language words and enhance translation speed and coverage.

VI. DISCUSSION

The proposed framework presents the conceptualization and development of a crowdsourcing-based editorial manager for developing and evaluating sign language corpus. It ensures that the volunteer-submitted gestures are of sufficient quality and can be uploaded to a digital repository. It also demonstrates that crowdsourcing can potentially ensure sustainable development, maintenance, extension, and to an extend standardization of a sign language and its various regional dialects. Similarly, it also shows the potential of wider public engagement for the deaf community.

The whole data collection process while involving the crowd was based on the guidelines provided for crowdsourcing-based natural language processing tasks [28]. It involves *a*) task design; *b*) data acquisition process; *c*) execution and monitoring; and *d*) data aggregation and evaluation. Whereby, the whole activity was well-defined and decomposed into several tasks including user registration, gesture submission, editorial process, conversion of a gesture into sign writing notation. Subsequently, the crowd was involved in gesture submission and validation. Later on, in stage-II data was prepared and collected with the help of the crowd. The project was executed in stage-III with the help of involved stakeholders the deaf community, language experts, and the technical support team. Here, stakeholder was taken onboard and trained to use the system to manage all the tasks. Lastly, data evaluation and aggregation were conducted, where the human videos submitted by the deaf contributors were evaluated by the language experts during the editorial process. While the quality of conversion into sign writing notation was evaluated by converting the sign writing notation into an avatar and getting it evaluated by the deaf community.

TABLE 12. Comparison with existing corpora for natural to sign language translation [19].

Corpus	Language Pairs	Sentences
Othman and Jemni	English-ASL	431
Bonham	English-ASL	11,000
Cate and Hussian	English-ASL	600
-	Italian-Italian SL	585
RWTH-phoenix	German-German SL	2486
-	Chinese-Taiwanese SL	1983
-	Catalan-Catalan SL	199
ATIS	English-iris SL	595
ATIS	German-German SL	595
ATIS	English south African SL	595
ID	Spanish-Spanish SL	2000
DL	Spanish-Spanish SL	2000
eSign-3D	Spanish-Spanish SL	5000
This Corpus	**English/Urdu-PSL**	**8300 and 700+ gestures**

It is pertinent to compare the collected data with the existing data-sets developed for natural to sign language translation. Table 12 presents the data-sets developed for the evaluation of natural to sign language translation. It can be clearly seen that most of the data-sets comprise of less than 2000 sentences. There is only 1 data-set for English to ASL translation that comprises of 11,000 sentences. Our data-set comprises of more than 700 gestures in a machine readable format, and more than 8000 sentences, and with a variety of sentences with respect to tenses and meanings.

The involved experts and the deaf community expressed a great deal of excitement while using this system. However, they also provided useful feedback for further expansion, scalability, and other improvements for the proposed framework. They wanted it to be advertised at a good scale so that the deaf community of the whole country should be involved in this initiative. Most of the community members were satisfied with the quality of the review process, but some of them showed concerns over the review process. They also pointed out that a large number of reviewers and editors need to be on board to scale up the system at the country level. Most of the contributors and language experts called the call for gestures as an exciting platform for coming up with new gestures for modern terms. The existing system is very bureaucratic and slow and is almost a manual system, which makes it very slow. Similarly, the inclusion of different dialects was well accepted and appreciated by the language experts as well as the deaf community. The major reason behind this was the fact that for PSL not many gestures have been officially standardized, and different regions claim their versions to be standard gestures. While the provision of dialect in the gesture submission system has provided everyone a chance to submit their version of the gesture under the appropriate dialect and region. Lastly, most of the involved contributors and language experts wanted to work consistently with this system for a longer period of time.

A. POTENTIAL BENEFITS

The proposed crowdsourcing-based framework to develop sign language corpus for different dialects can potentially lead to a variety of benefits. It can help to develop a huge repository of verified and validated gestures of different regions and dialects of a sign language. The conversion of these gestures into machine-readable format can open ways to develop domain-specific and generic machine translation applications for the deaf community. These applications can be flexible and can be multi-dialect translation applications, where the user can choose the dialect in which the gesture should be played after translation. Apart from this, the machine-readable version takes very small space in the memory and can help developing applications for portable devices like cell phones etc. Another interesting benefit from this crowdsourcing platform is that different gestures of the same word in the same dialect can be compiled into a single data-set that can help as a real data-set for gesture recognition systems for that sign language. Lastly, the sentence base corpus can be used for training and testing sign language translation models. Furthermore, it can also help standardization of gestures of sign language. Similarly, it can be used as a platform to design sign language gestures for new terminologies and words by engaging language experts and the deaf community.

VII. CONCLUSION

This article presents a framework that involves a process for developing a two-level corpus for different dialects of a sign language, namely, word and gesture level parallel corpus; and sentence level translation corpus. The corpus building process involves the deaf community as a contributor in the form of the crowd. Apart from the deaf community, it also involves the language experts to evaluate the data submitted by the deaf community through an editorial process. Furthermore, a technical team also converts the collected gestures into a machine-readable sign writing notation that can help to generate an avatar performing the same gesture. The theoretical framework has been developed into a data acquisition portal which was used by more than 50 people in different capacities to collect few hundred gestures. While, the technical team converted these gestures into a machine-readable format, and subsequently generated equivalent avatars from them. As a whole, almost a word-level corpus of almost 700 distinct words was created, while a sentence-level translation corpus comprises more than 8000 pairs of source and target sentences.

The whole process and the output avatar were evaluated by the deaf subjects. Almost all the involved deaf subjects appreciated this framework and considered it necessary for the development of sign languages. They were also briefed about the potential benefits of this framework, and considering them they were eager to consistently contribute to this framework in the capacity of contributor and validator, to add new gestures and validate and rate the avatars, respectively.

This research can be extended in many different ways. Firstly, there is a need to scale this concept up by engaging a large number of people as contributors and also need to engage many more language experts to expand it into a fully working system. Secondly, it can be expanded as a global sign language corpus by involving subjects and language experts from different countries, thus extending it into a global sign language corpus. Similarly, the linguistic information about the language can be collected with the help of the involved deaf community. Lastly, a multilingual sign language translation system can also be developed and tested using the multilingual sentence level corpus.

REFERENCES

[1] L. Zhao, K. Kipper, W. Schuler, C. Vogler, N. Badler, and M. Palmer, "A machine translation system from English to American Sign Language," in *Proc. Conf. Assoc. Mach. Transl. Amer.* Berlin, Germany: Springer, 2000, pp. 54–67.

[2] I. Marshall and É. Sáfár, "A prototype text to british sign language (BSL) translation system," in *Proc. 41st Annu. Meeting Assoc. Comput. Linguistics (ACL)*, 2003.

[3] J. Bungeroth and H. Ney, "Statistical sign language translation," in *Proc. Workshop Represent. Process. Sign Lang. (LREC)*, vol. 4, May 2004, pp. 105–108.

[4] R. San-Segundo, V. López, R. Martín, D. Sánchez, and A. GarcíA, "Language resources for Spanish-Spanish sign language (LSE) translation," in *Proc. 4th Workshop Represent. Process. Sign Lang., Corpora Sign Lang. Technol. (LREC)*, 2010, pp. 208–211.

[5] L. Van Zijl and A. Combrink, "The South African sign language machine translation project: Issues on non-manual sign generation," in *Proc. Annu. Res. Conf. South Afr. Inst. Comput. Scientists Inf. Technologists IT Res. Developing Countries*, 2006, pp. 127–134.

[6] V. K. Verma and S. Srivastava, "Toward machine translation linguistic issues of Indian sign language," in *Speech and Language Processing for Human-Machine Communications*. Singapore: Springer, 2018, pp. 129–135.

[7] T. B. D. Nguyen, T. N. Phung, and T. T. Vu, "A rule-based method for text shortening in Vietnamese sign language translation," in *Information Systems Design and Intelligent Applications*. Singapore: Springer, 2018, pp. 655–662.

[8] F. Yasir, P. W. C. Prasad, A. Alsadoon, A. Elchouemi, and S. Sreedharan, "Bangla sign language recognition using convolutional neural network," in *Proc. Int. Conf. Intell. Comput., Instrum. Control Technol. (ICICICT)*, Jul. 2017, pp. 49–53.

[9] N. Sabir, A. Shahzada, S. Atta, A. Abid, Y. Daniaal, S. Farooq, T. Mushtaq, and I. Khan, "A vision based approach for Pakistan sign language alphabets recognition," *Pensee J.*, vol. 76, no. 3, pp. 274–285, 2014.

[10] J. Tumsri and W. Kimpan, "Thai sign language translation using leap motion controller," in *Proc. Int. Multiconf. Eng. Comput. Scientists*, 2017, pp. 46–51.

[11] H. Luqman and S. A. Mahmoud, "Automatic translation of arabic text-to-Arabic sign language," *Universal Access Inf. Soc.*, vol. 18, no. 4, pp. 939–951, Nov. 2019.

[12] M. Karbasi, A. Zabidi, I. M. Yassin, A. Waqas, and Z. Bhatti, "Malaysian sign language dataset for automatic sign language recognition system," *J. Fundam. Appl. Sci.*, vol. 9, no. 4S, pp. 459–474, 2017.

[13] N. S. Khan, A. Abid, K. Abid, U. Farooq, M. S. Farooq, and H. Jameel, "Speak Pakistan: Challenges in developing Pakistan sign language using information technology," *South Asian Stud.*, vol. 30, no. 2, p. 367, 2015.

[14] K. Abid, N. S. Khan, U. Farooq, M. S. Farooq, M. A. Naeem, and A. Abid, "A roadmap to elevate Pakistan sign language among regional sign languages," *South Asian Stud.*, vol. 33, no. 2, pp. 461–473, 2018.

[15] J. Hutchinson, "Literature review: Analysis of sign language notations for parsing in machine translation of SASL," Rhodes Univ., Grahamstown, South Africa, 2012.

[16] K. Cormier, J. Fenlon, T. Johnston, R. Rentelis, A. Schembri, K. Rowley, R. Adam, and B. Woll, "From corpus to lexical database to online dictionary: Issues in annotation of the BSL Corpus and the development of BSL SignBank," in *Proc. 5th Workshop Represent. Process. Sign Lang., Interact. Between Corpus Lexicon, Int. Conf. Lang. Resour. Eval. (LREC)*, 2012, pp. 7–12.

[17] E. Taljard and D. Prinsloo, "African language dictionaries for children-A neglected genre," *Lexikos*, vol. 29, no. 1, pp. 199–223, 2019.

[18] J. L. Fuertes, Á. L. González, G. Mariscal, and C. Ruiz, "Bilingual sign language dictionary," in *Proc. Int. Conf. Comput. Handicapped Persons*. Berlin, Germany: Springer, Jul. 2006, pp. 599–606.

[19] J. Porta, F. López-Colino, J. Tejedor, and J. Colás, "A rule-based translation from written spanish to spanish sign language glosses," *Comput. Speech Lang.*, vol. 28, no. 3, pp. 788–811, May 2014.

[20] E. Suzuki, T. Suzuki, and K. Kakihana, "On the Web trilingual sign language dictionary to learn the foreign sign language without learning a target spoken language," in *Proc. LREC*, 2006, pp. 2307–2310.

[21] J. H. Kristoffersen, T. Troelsgård, *The Electronic Lexicographical Treatment of Sign Languages: The Danish Sign Language Dictionary.* Oxford, U.K.: Oxford Univ. Press, 2012.

[22] T. Hanke, L. König, S. Wagner, and S. Matthes, "DGS corpus & dicta-sign: The hamburg studio setup," in *4th Workshop Represent. Process. Sign Languages: Corpora Sign Lang. Technol. (CSLT)*, Valletta, Malta, 2010, pp. 106–110.

[23] J. Forster, C. Schmidt, O. Koller, M. Bellgardt, and H. Ney, "Extensions of the sign language recognition and translation corpus RWTH-phoenix-weather," in *Proc. LREC*, 2014, pp. 1911–1916.

[24] Y.-H. Chiu, C.-H. Wu, H.-Y. Su, and C.-J. Cheng, "Joint optimization of word alignment and epenthesis generation for chinese to taiwanese sign synthesis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 28–39, Jan. 2007.

[25] (2020). *National Association of the Deaf–NAD*. Accessed: Dec. 13, 2020. [Online]. Available: https://www.nad.org/

[26] (2020). *Directorate General of Special Education, Islamabad, Pakistan*. Accessed: Dec. 13, 2020. [Online]. Available: http:// dgse.gov.pk/ html/index.html

[27] (2020). *National Association of the Deaf, India*. Accessed: Dec. 12, 2020. [Online]. Available: http://nadindia.org/

[28] M. Sabou, K. Bontcheva, L. Derczynski, and A. Scharl, "Corpus Annotation through Crowdsourcing: Towards Best Practice Guidelines," in *Proc. LREC*, May 2014, pp. 859–866.

[29] K. N. Riemer, T. Björkstrand, J. Mesch, and C. Börstell, "Crowdsourcing for the Swedish sign language dictionary," in *Proc. 8th Workshop Represent. Process. Sign Lang.*, Miyazaki, Japan, May 2018 pp. 171–174.

[30] U. Farooq, M. Shafry M. Rahim, A. Abid, and N. S. Khan, "A process to develop Pakistan sign language corpus using crowd sourcing," *Int. J. Innov. Technol. Exploring Eng.*, vol. 2020, pp. 2098–2103, Mar. 2020.

[31] M. Russell, M. Kavanaugh, J. Masters, J. Higgins, and T. Hoffmann, "Computer-based signing accommodations: Comparing a recorded human with an avatar," *J. Appl. Test. Technol.*, vol. 10, no. 3, pp. 1–20, 2014.

[32] Z. Kang, "Spoken language to sign language translation system based on HamNoSys," in *Proc. Int. Symp. Signal Process. Syst. (SSPS)*, 2019, pp. 159–164.

[33] Y. Bouzid and M. Jemni, "An animated avatar to interpret signwriting transcription," in *Proc. Int. Conf. Electr. Eng. Softw. Appl.*, Mar. 2013, pp. 1–5.

[34] M. Borg and K. P. Camilleri, "Towards a transcription system of sign language video resources via motion trajectory factorisation," in *Proc. ACM Symp. Document Eng.*, Aug. 2017, pp. 163–172.

[35] E. Efthimiou, S.-E. Fotinea, T. Goulas, A. Vacalopoulou, K. Vasilaki, and A.-L. Dimou, "Sign language technologies in view of future Internet accessibility services," in *Proc. 11th Pervas. Technol. Rel. Assistive Environments Conf.*, Jun. 2018, pp. 495–501.

[36] I. A. Chiriac, L. Stoicu-Tivadar, and E. Podoleanu, "Comparing video and avatar technology for a health education application for deaf people," in *MIE*, May 2015, pp. 516–520.

[37] T. Causer, J. Tonra, and V. Wallace, "Transcription maximized; expense minimized? Crowdsourcing and editing *The Collected Works of Jeremy Bentham*," *Literary Linguistic Comput.*, vol. 27, no. 2, pp. 119–137, 2012.

[38] A. Matsunaga, A. Mast, and J. A. B. Fortes, "Workforce-efficient consensus in crowdsourced transcription of biocollections information," *Future Gener. Comput. Syst.*, vol. 56, pp. 526–536, Mar. 2016.

[39] M. M. Terras, *Crowdsourcing in the Digital Humanities*. Hoboken, NJ, USA: Wiley-Blackwell, 2016.

[40] P. Ruiz, "Concept-based and relation-based corpus navigation: Applications of natural language processing in digital humanities," Doctoral dissertation, PSL Res. Univ., Paris, France, 2017.

[41] N. S. Khan, A. Abid, and K. Abid, "A novel natural language processing (NLP)–based machine translation model for English to Pakistan sign language translation," *Cogn. Comput.*, vol. 12, pp. 748–765, May 2020.

[42] A. Becker, F. Kepler, and S. Candeias, "A Web tool for building parallel corpora of spoken and sign languages," in *Proc. 10th Int. Conf. Lang. Resour. Eval. (LREC)*, May 2016, pp. 1438–1445.

[43] D. Bragg, O. Koller, M. Bellard, L. Berke, P. Boudreault, A. Braffort, and R. M. Morris, "Sign language recognition, generation, and translation: An interdisciplinary perspective," in *Proc. 21st Int. ACM SIGACCESS Conf. Comput. Accessibility*, Oct. 2019, pp. 16–31.

[44] (2020). *Deafness and Hearing Loss*. Accessed: Apr. 21, 2020. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

[45] A. Bayazed, O. Torabah, R. AlSulami, D. Alahmadi, A. Babour, and K. Saeedi, "SDCT: Multi-dialects corpus classification for Saudi Tweets," *Methodology*, vol. 11, no. 11, 2020.

[46] D. Alahmadi, A. Babour, K. Saeedi, and A. Visvizi, "Ensuring inclusion and diversity in research and research output: A case for a language-sensitive NLP crowdsourcing platform," *Appl. Sci.*, vol. 10, no. 18, p. 6216, Sep. 2020, doi: 10.3390/app10186216.

[47] M.-H. Wu and A. Quinn, "Confusing the crowd: Task instruction quality on Amazon mechanical turk," in *Proc. AAAI Conf. Hum. Comput. Crowdsourcing*, vol. 5, no. 1, 2017, pp. 206–215. [Online]. Available: https://ojs.aaai.org/index.php/HCOMP/article/view/13317

[48] B. Shmueli, J. Fell, S. Ray, and L.-W. Ku, "Beyond fair pay: Ethical implications of NLP crowdsourcing," 2021, *arXiv:2104.10097*. [Online]. Available: http://arxiv.org/abs/2104.10097

[49] U. Farooq, M. S. M. Rahim, N. Sabir, A. Hussain, and A. Abid, "Advances in machine translation for sign language: Approaches, limitations, and challenges," *Neural Comput. Appl.*, to be published, doi: 10.1007/s00521-021-06079-3.

[50] U. Farooq, A. Asmat, M. S. B. M. Rahim, N. S. Khan, and A. Abid, "A comparison of hardware based approaches for sign language gesture recognition systems," in *Proc. Int. Conf. Innov. Comput. (ICIC)*, Nov. 2019, pp. 1–6.

UZMA FAROOQ received the B.S. degree in computer science from the University of the Punjab, Pakistan, in 2005, and the M.S. degree in computer science from the National University of Computer and Emerging Sciences, Pakistan, in 2007. She is currently pursuing the Ph.D. degree with Universiti Teknologi Malaysia, Malaysia. She is currently working as an Assistant Professor with the Department of Software Engineering, University of Management and Technology, Pakistan. Her research interests include assistive technologies, machine learning, and natural language processing.

MOHD SHAFRY MOHD RAHIM was born in Malaysia. He received the bachelor's degree in computer science and the M.S. and Ph.D. degrees in computing from UTM, Malaysia. He is currently a Professor with UTM, where he is leading a research group comprising of faculty members and graduate students. He is also the Chair of undergraduate students at UTM. He has published research article in several renowned international journals and conference papers. He has appeared as a key note speaker for many international conferences.

NABEEL SABIR KHAN was born in Lahore, Pakistan. He received the M.C.S. and M.S. degrees from the University of Central Punjab, and the Ph.D. degree from the University of Management and Technology, Pakistan, in 2020. He is currently working as an Assistant Professor with the University of Management and Technology. He has more than 13 years of teaching experience. He is also the Regional Director of ACM-ICPC ASIA, Lahore. His research interests include theory of programming language, machine translation, and computer science education.

SAIM RASHEED was born in Pakistan. He received the Ph.D. degree in brain computer interaction from the University of Milan, Italy, in 2011. He has almost 20 years of experience in academia and research. He has been actively participating and contributing in many international conferences and journals. His current research interests include human–computer interaction, gamification, and natural language processing.

ADNAN ABID (Member, IEEE) received the B.S. degree from the National University of Computer and Emerging Science, Pakistan, in 2001, the M.S. degree in information technology from the National University of Science and Technology, Pakistan, in 2007, and the Ph.D. degree in computer science from the Politecnico di Milano, Italy, in 2012. He spent one year at EPFL, Switzerland, to complete his M.S. thesis. He is currently working as a Professor with the Department of Computer Science, University of Management and Technology, Pakistan. He has published more than 80 articles in different international journal and conference papers. His research interests include computer science education, information retrieval, and data management. He is a member of the IEEE Education Society. He has served as a reviewer for many international conference papers and journal articles. He is an Associate Editor of IEEE ACCESS journal and an Editor of *PeerJ Computer Science*.