

Jumla Sign Language Annotation Tool: an overview

Achraf Othman, Oussama El Ghoul
Mada Center, Qatar

Abstract –

This paper describes an ongoing project on developing a new web-based tool for annotating Sign languages. This tool is used to annotate the First Qatari Sign Language dataset called “Jumla Dataset: The Jumla Qatari Sign Language Corpus” with written Arabic text. The annotation of videos in Qatari Sign Language (QSL) takes input from signers to identify the Arabic glosses components toward representing the QSL in a written way with high accuracy, furthermore to the use of the annotation output in the development of computational Sign Language tools. The QSL annotation is based on an input of 4 videos recorded by deaf persons or Sign Language interpreters from different angles (front, left side, right side, and facial view). The output is a JSON file containing all the interpreted sentences given as an entry record. The glosses are annotated for each period and aligned with the Arabic content. Moreover, the presented tool, available as an open-source, provides a management system to classify all records from cameras, motion capture systems, and edited files in addition to the possibility to create components for each gloss annotation terminology depending on the target Sign Language.

Keywords: Annotation Tool, Sign Language Corpus, Qatari Sign Language, Gloss Annotation

Introduction

Sign languages present the primary way of communication in persons with hearing impairment and the deaf communities used in large groups of persons worldwide. They are considered natural visual languages different than spoken languages [1]. They have a rich grammatical structure with unique syntax and vocabulary [2] compared to spoken languages [3]. However, without a written form, the information cannot be processed by machine and interpreted by humans. For this, sign language can be written in a way that permits anyone to use it to build systems and to develop applications. Sign languages have a spatiotemporal structure different from the linear structure of spoken language. This structure challenges developing the written form to cover syntactic and semantic levels. Like spoken languages, sign language also has different grammatical levels that require annotation for processing tasks. For each, it can be done automatically or manually according to research progress.

The annotation of sign language allows writing all linguistic information related to the signed language using a text description. Annotation is defined as a special kind of task where we add additional information to a piece of text for its better interpretation, analysis, comprehension,

and application [4]. In this paper, we present a new web-based tool developed by Mada to annotate SL. The Jumla Sign Language project is supported by the Mada Innovation Program (MIP) [5]. It provides the first Qatari Sign Language avatar to translate Arabic text to Qatari and Arabic Sign Language.

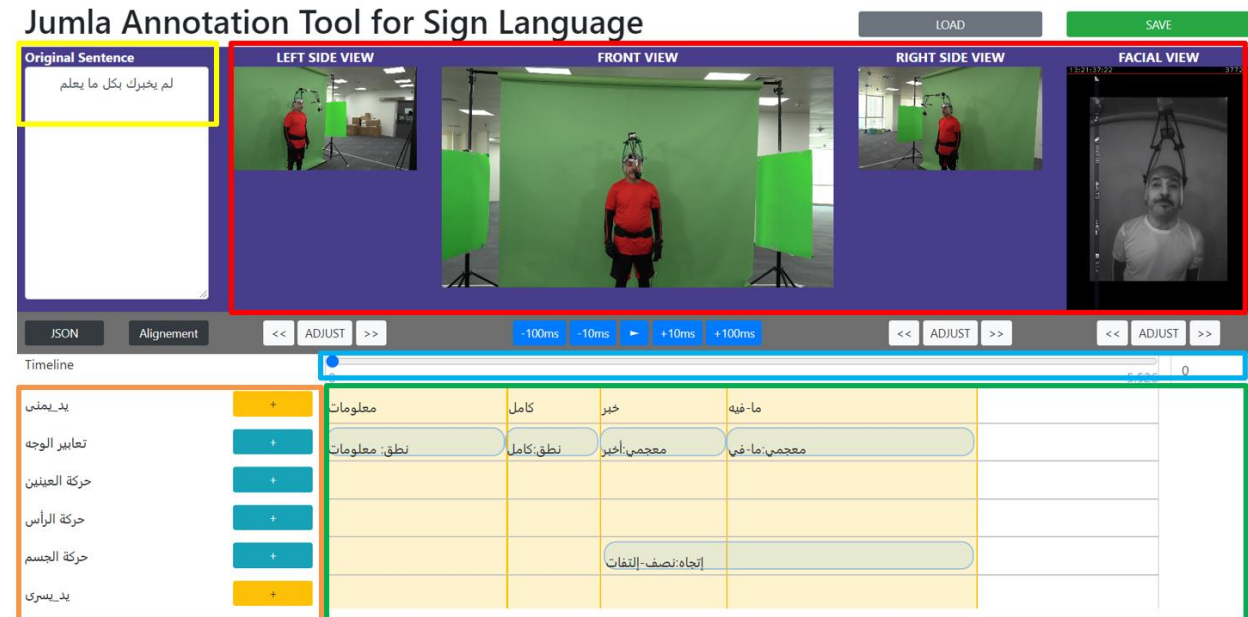


Figure 1. Jumla annotation tool for sign language.

Literature review

In sign language, the annotation is defined as a critical or explanatory commentary or analysis, comment added to a text, the process of writing such comment or commentary, in computing, metadata added to a document or program. Compared to notation which is defined as a system of characters, symbols, or abbreviated expressions used in an art or science or in mathematics or logic to express technical facts or quantities [6]. In computational sign language, we will refer to the term annotation to describe the written form of sign language. The term notation will be used in linguistics used in the works of Stokoe and Sutton [7].

For the annotation systems used to describe signs in recorded videos, several works can be classified into three categories:

- (1) Manual annotation systems: the user must add comments and descriptions for each sign
- (2) Automatic annotation: that uses algorithms and machine learning to annotate a video automatically. [Until today, there is no accurate, high-performance system able to reach the same level of manual annotation systems
- (3) Supervised annotation systems: they are hybrid systems that allow the user to add manual labels and tags with automatic scripts.

For any type of annotation system, the following sign language components must be included during adding the labels and notes: (1) manual signs: hand shapes, hand orientation, shoulder movements, body movements, body orientation, head movements, eye gaze, facial expression (eyebrows, eyelids, and nose), mouthing (mouth pictures and mouth gestures) and emotion if applicable. It can also have extra grammatical labels like part-of-speech tagging and lemmatization. One of the most cited tools is ELAN [8]. It is a free, multimodal annotation tool for digital audio and video media. It supports multileveled transcription of up to six synchronized video files per annotation document [9]. It was used to annotate the Corpus NGT is a collection of almost 72 hours of dialogues of 92 different signers for whom NGT is the first language [10]. After releasing the software, several research projects used the ELAN tool to annotate videos in different sign languages.

Objectives

The main goal of this work is to provide a free, open-source tool that helps researchers and linguistics annotate videos in different sign languages. Not limited to this and from a research perspective, through the present tool, we try to solve the following issues related to sign language annotation tasks and processing:

- (1) Development of an online and collaborative platform and interface with voting systems
- (2) Inclusion of manual and non-manual with the integration of manual and automatic labeling techniques (supervised, semi-supervised, and non-supervised)
- (3) Integration of multi-level annotation layers. For a sign language input, we can extend the annotation to an intra-linguistic level and an extra-linguistic level.

After recording the raw video, we proceed with tasks to build the corpora and annotate the video. The pre-processing task is essential to crop the video and make it ready for annotation. The annotation task can be performed without the processing phase. The annotation output can enhance the quality of the processing phase output. The meta-data phase enables the annotator to define the types of labels used during the annotation task and can be the output of the pre-processing phase. The meta-data types can be grammatical labels, or any information used during the annotation phases, which helps in the processing phase too. The corpora management model inspires the management model by Dash [4]. We adopt the methodology described in the work of Ide and Suderman, “The Linguistic Annotation Framework” [11], developed by the International Standards Organization (ISO) as a standard ISO TC37 SC4. It identifies the fundamental properties and principles for representing linguistic annotations and leads to the design of an abstract data model. The project is ongoing and still did not implement the complete process of the framework. Until today, an online tool was developed to support annotators to annotate the sign language videos. This task is done as follows:

- (1) The webmaster uploads the video to the tool with the sentence in written form
- (2) A deaf person accesses the video and segments the video in sign. Each sign will start in a specific frame and end in another frame

- (3) A second deaf person add labels related to non-manual signs like facial expression, eye gaze, head movement, body movement, and non-dominant hand
- (4) a sign language expert reviews the complete annotation and confirm it
- (5) a workshop is organized to review all annotations by a committee and approve the proposed annotation.

The interface contains the following components (Fig 1):

- **Yellow box:** the sentence in original language (for example in the Arabic language: لم يخبرك بكل ما يعلم (in English: He did not tell you everything he knows))
- **Red box:** contains a synchronized preview of 4 recorded videos (front view, left view, right view, and facial view). The annotator can use only one video (for example, front view). In our case, we used four views to ensure annotation accuracy.
- **Blue box:** the timeline with a progress bar.
- **Orange box:** contains a list of the sign language components. Each component is added in a separate layer. The tool admin can manage the sign language components (adding, removing) because sign language differs.
- **Green box:** the annotation zone that contains all labels added by the annotators will be shown from a start frame to an end frame.

For example sentence, we can observe that the sentence was not signed word by word but differently: “ما فيه | خبر | كامل | معلومات” (in English: information | full | news | no). Moreover, to pronunciation labels, non-manual signs to complete the whole meaning of the sentence and the body movement used to place the “no” sign in another space position for reference.

Annotation tools for sign language must follow a written form. For the Arabic language, we can follow the gloss annotation methodology as applied in American Sign Language [13]. There are several advantages in annotating sign language:

Through sign language annotation tasks, we can understand the structure of sign language sentences and linguistic structure in the absence of sign language grammar.

- Through the annotation tasks, we can identify and retrieve prosody, pronunciation, grammar, meaning, sentence, syntactic, semantic, discourse, and figurative information from the sign language video.
- We can identify the time, location of objects, subjects, and the size of the signing space.
- We can provide a meta-data model for processing purposes.

Moreover, annotation in sign language opened the doors to new research topics, which are sign language segmentation and automatic annotation and not limited to [12] [13]. The benefits of annotated sign language cover several domains such as technology, the commercial sector for

machine translation and sign language recognition, ethnographic and humanistic domains. Furthermore, thanks to the fast growth of technology and artificial intelligence, the more we have annotated data, the more research can produce new directions in improving the quality of life and independent living for deaf persons [14].

Conclusion

Sign language annotation tools solve several issues faced in sign language processing from linguistic problems, automatic sign language recognition systems, and the development of statistical machine translation systems. In this work, we completed the annotation of 1600 prerecorded sentences provided for the public. The annotation task takes time to ensure the quality of the output. Moreover, we presented a novel annotation tool for sign language covering several levels of annotation tasks from intra-linguistic to extra-linguistic. The output of the annotation tasks is critical to progress research in computational sign language in general. In addition, this work presents an original contribution to Qatari sign language processing and can be extended to other Arabic and foreign sign languages. Moreover, this can be very useful to enhance and upgrade educational skills within deaf students in primary education.

Acknowledgment

We highly appreciate and are grateful for the team of the Mada Innovation Program and contributors from the Qatari deaf community, The Qatari Center of Social Cultural for the Deaf and Audio Education Complex. Mocap cleaning tasks and the Jumla sign language project have seen the light thanks to their support and contribution.

References

- [1] J. G. Kyle, J. Kyle, B. Woll, G. Pullen, and F. Maddix, *Sign language: The study of deaf people and their language*. Cambridge university press, 1988.
- [2] K. Emmorey, *Language, cognition, and the brain: Insights from sign language research*. Psychology Press, 2001.
- [3] U. Bellugi and S. Fischer, "A comparison of sign language and spoken language," *Cognition*, vol. 1, no. 2–3, pp. 173–200, 1972.
- [4] N. S. Dash, *Language Corpora Annotation and Processing*. Springer, 2021.
- [5] D. Al Thani, A. Al Tamimi, A. Othman, A. Habib, A. Lahiri, and S. Ahmed, "Mada Innovation Program: A Go-to-Market ecosystem for Arabic Accessibility Solutions," in *2019 7th International conference on ICT & Accessibility (ICTA)*, 2019, pp. 1–3.
- [6] "Annotation vs Notation - What's the difference?," *WikiDiff*, May 05, 2015. [//wikidiff.com/annotation/notation](https://wikidiff.com/annotation/notation) (accessed Nov. 30, 2021).
- [7] M. Kato, "A study of notation and sign writing systems for the deaf," *Intercult. Commun. Stud.*, vol. 17, no. 4, pp. 97–114, 2008.

- [8] O. Crasborn and H. Sloetjes, "Enhanced ELAN functionality for sign language corpora," in *6th International Conference on Language Resources and Evaluation (LREC 2008)/3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, 2008, pp. 39–43.
- [9] O. A. Crasborn and H. Sloetjes, "Using ELAN for annotating sign language corpora in a team setting," 2010.
- [10] O. A. Crasborn and I. E. P. Zwitserlood, "The Corpus NGT: an online corpus for professionals and laymen," 2008.
- [11] N. Ide and K. Suderman, "The Linguistic Annotation Framework: a standard for annotation interchange and merging," *Lang. Resour. Eval.*, vol. 48, no. 3, pp. 395–418, 2014.
- [12] H. Chaaban, M. Gouiffès, and A. Braffort, "Automatic Annotation and Segmentation of Sign Language Videos: Base-level Features and Lexical Signs Classification.," in *VISIGRAPP (5: VISAPP)*, 2021, pp. 484–491.
- [13] A. Othman and M. Jemni, "Designing high accuracy statistical machine translation for sign language using parallel corpus: case study English and American Sign Language," *J. Inf. Technol. Res. JITR*, vol. 12, no. 2, pp. 134–158, 2019.
- [14] A. Othman and M. Jemni, "Statistical sign language machine translation: from English written text to American sign language gloss," *ArXiv Prepr. ArXiv11120168*, 2011.