Proceedings of the 9th Workshop on the Representation and Processing of Sign Languages, pages 117–122
Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020

Towards Kurdish Text to Sign Translation

Zina Kamal, Hossein Hassani
University of Kurdistan Hewlêr, University of Kurdistan Hewlêr
Kurdistan Region - Iraq, Kurdistan Region - Iraq
{z.kamal3, hosseinh}@ukh.edu.krd

Abstract

The resources and technologies for sign language processing of resourceful languages are emerging, while the low-resource languages are falling behind. Kurdish is a multi-dialect language, and it is considered a low-resource language. It is spoken by approximately 30 million people in several countries, which denotes that it has a large community with hearing-impairments as well. This paper reports on a project which aims to develop the necessary data and tools to process the sign language for Sorani as one of the spoken Kurdish dialects. We present the results of developing a dataset in HamNoSys and its corresponding SiGML form for the Kurdish Sign lexicon. We use this dataset to implement a sign-supported Kurdish tool to check the accuracy of the sign lexicon. We tested the tool by presenting it to hearing-impaired individuals. The experiment showed that 100% of the translated letters were understandable by a hearing-impaired person. The percentages were 65% for isolated words, and approximately 30% for the words in sentences. The data is publicly available at `https://github.com/KurdishBLARK/KurdishSignLanguage` for non-commercial use under the CC BY-NC-SA 4.0 licence.

Keywords: Sign Language, Kurdish Language Processing, Kurdish to Sign, HamNoSys, SiGML

1. Introduction

The studies on sign language processing have been emerging, but many areas are still unexplored (Cormier et al., 2019). As might be expected, this area of research has even not been initiated yet for many under-resourced languages.

Kurdish, a multi-dialect language which is spoken by approximately 30 million people in different countries, is considered an under-resourced language (Hassani, 2018). It is also written in different scripts. The usage of the scripts changes according to the geographical situation (Hassani and Medjedovic, 2016).

The current literature does not report on visible research on Kurdish Sign Language (KuSL) processing, nor are there any publicly available resources for this topic. This research focuses on text to sign conversion for the Sorani dialect of Kurdish.

Sign language is the main communication method among the hearing-impaired community. This language is based on visual interaction rather than using sound. The interactions happen by manual and non-manual signs and finger spelling (Cooper et al., 2011). Hand and body movement, shape, orientation and location are within manual signs (Kelly et al., 2009), while facial expressions, eye gaze, and shoulder movement are called non-manual signs (Halawani, 2008). Furthermore, the finger spelling is used to spell letters of certain words, for example, names and technical terms that do not have sign equivalents (Liwicki and Everingham, 2009).

Normally, the communication between two hearing-impaired persons is smooth and understandable. The real challenge begins when a hearing person wants to interact with a hearing-impaired person (Wazalwar and Shrawankar, 2017).

Generally, if the target hearing-impaired person is educated, they try to communicate by exchanging written texts. Otherwise, they turn to a human sign language interpreter as a recourse if available, or else perhaps they end up with serious miscommunication (Wazalwar and Shrawankar, 2017).

Although the spoken Kurdish dialects use different lexicons (Ahmadi et al., 2019), the Kurdish Sign language, which is used in the Kurdistan Region of Iraq (KRI), uses the same lexicon among the hearing-impaired community regardless of the spoken dialect. While according to Jepsen et al. (2015) KuSL is not standardized, applying guidelines by Mohammed (2007) and using the Kurdish Sign dictionaries (Nashat Salim et al., 2013; Ghazi Dizayee, 2000) in the KRI education programs show some efforts towards KuSL standardization.

We develop a Kurdish Sign lexicon using the Kurdish Sign Language Dictionary (KuSLD) (Ghazi Dizayee, 2000), which is used in KRI.

Currently, no Kurdish Sign corpus is available, hence we aim at making Sorani texts sign-supported. That is, in the text conversion process we follow the spoken language and not the sign language structure. Sorani texts are mostly written in Persian-Arabic script (Hassani, 2018) hence we use the developed Kurdish Sign lexicon to make this type of the Sorani texts sign-supported.

The rest of this paper is organized as follows. Section 2. provides a brief background on sign language processing, Section 3. reviews the related work, Section 4. presents our approach, Section 5. illustrates the developed dataset, Section 6. discusses the results, finally, Section 7. concludes the paper.

2. Sign Language Processing

Sign languages are considered as genuine languages that place them among the minority languages (Senghas and Monaghan, 2002). Since sign languages consist

of visual gestures rather than voice as it is in spoken languages. The analysis and feature extraction of the former significantly differ from latter languages. However, for some languages, a variant of sign language also exists that follows the spoken/written language grammar, which is called sign-supported language (Elliott et al., 2008). The development of this variant is less challenging in the absence of required sign corpora and language models. The outcome could be used in various experimental and real-life occasions.

Several approaches exist to process sign languages. In the following sections, we discuss those approaches which are more related to our current stage of research.

2.1. Notation Systems

The sign visual gestures are normally denoted by special notations in order to be able to process them. Different notation systems are used to capture these gestures. The most popular ones are Stokoe, SignWriting, and HamNoSys.

Stokoe was one of the earliest attempts for a sign language notation system (McCarty, 2004). However, it was only concerned with manual sign representation, and it lacked any consideration for non-manual signs, such as eye gaze and shoulders movements, which are an essential entity to convey meaning by facial expression.

SignWriting represents the signed gestures spatially in a 2D canvas (Bouzid and Jemni, 2013b). It is designed to facilitate communication among the hearing-impaired community.

HamNoSys (Hamburg Notation System) is a phonetic translation system with iconicity, extensibility, and formal syntax characteristics used to denote sign languages (Hanke, 2004).

A comparative analysis by (Dhanjal and Singh, 2019) concluded that HamNoSys is the most widely used notation system for a variety of sign languages. HamNoSys symbols are available as a Unicode font (Hanke, 2004). This Unicode font symbolizes manual sign gestures and allows the generation of the signs by dividing the description into the handshapes, orientations, locations, and actions.

2.2. Markup Languages

To provide computer encoding for sign languages and to make their processing more efficient, several adoptions of the Extensive Markup Language (XML) have been suggested based on various sign notation systems.

The Sign Writing Markup Language (SWML) is a markup language proposed by da Rocha Costa and Dimuro (2001) based on SignWriting.

HamNoSys uses Signing Gesture Markup Language (SiGML), which gives a special XML tag to each HamNoSys symbol.

These markup languages are used in different applications, for instance, to be given to a 3D avatar to animate the signs.

3. Related Work

The only work on Kurdish Sign language processing that we were able to retrieve was by Hashim and Alizadeh (2018) wherein the researchers reported on their project on Kurdish Sign language recognition. That project focused on the recognition of Kurdish manual alphabets.

Therefore, as literature does not report on active studies on Kurdish Sign language processing, we review the topic in the context of other languages.

Sugandhi and Kaur (2018) introduced an online multilingual dictionary for avatar-based Indian Sign language. The system is designed to accept input from two languages English and Hindi. The input is transliterated into Hindi and then goes through the parser to be translated into Indian Sign Language (ISL). After extracting the root words of the input script, the target Hamburg Notations are retrieved from the database and converted into its corresponding SiGML. The generated SiGML is the input parameter for the Animation server, which uses Web Graphics Library (WebGL) for the avatar representation.

Aouiti (2013) proposed an approach to convert Arabic text into Arabic Sign language. The approach used an Arabic sentence/Sign language corpus as a core entity. The corpus includes Arabic sentences that were aligned with their corresponding sign representation. This helped to ensure that the represented sign refers to the real meaning of the input text. Afterward, the target sentence was syntactically and semantically analyzed by applying techniques, such as Morphological, Syntactic, Semantic, and Pragmatic analysis, which led to the generation of the glosses. The sign for each gloss was extracted from the corpus, which was sent to the avatar to be played.

Bouzid and Jemni (2013a) developed an avatar-based system to enhance the usability and readability of notation systems for deaf people. The system was developed using SignWriting (SW) notation and its markup language. Their focus was to make the path easier for hearing-impaired people to understand and represent signs in a written format. Since SW is presented in a 2D format and it is easy to guess the target gestures from the written notations, this helps hearing-impaired people to learn different sign languages depending on the SW notations. SW is designed for daily communication purposes rather than linguistic and corpus development and processing.

An automated reading system for SignWriting representation of Brazilian Sign language was introduced by Stiehl et al. (2015). They focused on SignWriting of several Brazilian signs and classified the symbols into several categories. Again, their purpose was to build a database of SignWriting representation for Brazilian Sign Language in order to involve hearing-impaired people into learning the notations and enable them to communicate with each other. This approach can also be used to have books, newspapers, dictionaries and such that are written in notation symbols and can be understood by hearing-impaired people or sign learn-

ers.

To summarize, we follow the approach of Sugandhi and Kaur (2018) because of two reasons. First, because HamNoSys is a proper method for corpus development, and second because SignWiriting is majorly used for the communication among the hearing-impaired community and not between the hearing-impaired community and people with no hearing difficulties.

4. Method

We develop a dataset based on KuSLD. We also prepare and adapt a tool to translate Sorani texts into Kurdish Sign language to be animated by an avatar.

We prepare the HamNoSys notations manually by analyzing the gestures available in KuSLD. We use the Ham2HPSG and eSIGN (Hanke and Popescu, 2003) to create the dataset and extract the SiGML codes.

We implement a sign-supported tool based on the architecture that is shown in Figure 1. In this architecture, the Language Model (LM), in its current form, is the developed dataset, which could be considered as the Kurdish (Sorani) sign lexicon. The input text goes through the tokenization process to extract the meaningful components from it.

Similar to the existing sign-supported tools for other languages, the translation is word-by-word for the words that are found in the LM. Otherwise, the word will be replaced by a sequence of its letters in the sign language. Then the whole text is compiled into SiGML files, which will be sent to an avatar to be animated.

We evaluate the tool by feeding it with input of four categories, namely alphabets, numbers, words, and sentences. The tool then plays the translation to the human individuals who are either hearing-impaired or Kurdish Sign language educators. As subjective understanding is not accurate (Kipp et al., 2011), we ask the testers to write down their understanding. We calculate the accuracy by the percentage of correctly understood cases for played alphabets, numbers, words, and sentences.

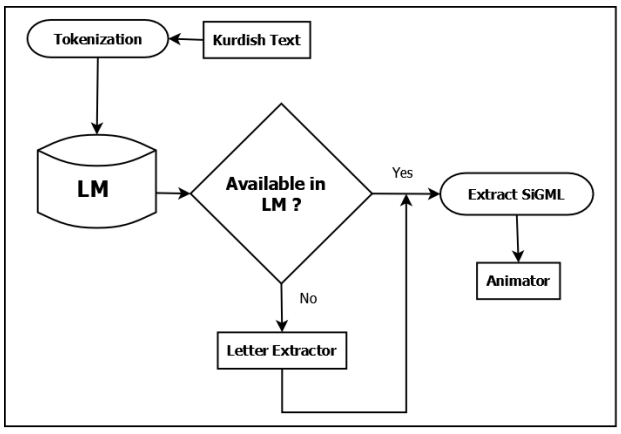

Figure 1: Kurdish text to sign proposed architecture

5. Developed Dataset

The KuSLD consists of 2315 different sign gestures from 38 different categories. Our dataset, currently, consists of 20% of each category. However, we converted the alphabet and numbers completely. This adds more entries to the dataset, which sums up to approximately 560 entries.

The KuSLD categories are listed in Table 1. A sample for the prepared HamNoSys for Kurdish letters and words is shown in Figures 2 and 3.

HamNoSys	Word
	من/Me
	سڵاو/Hello
	ئەمڕۆ/Today
	کتێب/Book
	کۆمپیوتەر/Computer
	خوێندن/Read

Figure 2: HamNoSys Sample for Kurdish words

HamNoSys	Letter
	ئ/E
	ا/A
	ب/B
	پ/P
	ت/T
	ج/C

Figure 3: HamNoSys Sample for Kurdish letters

We extracted the generated SiGML for the corresponding HamNoSys dataset, which was sent to an avatar to be animated. Two samples of extracted SiGML for the letter "ب" (B) and the word "زانکۆ" (University) in Kurdish are shown in Figures 4 and 5.

6. Findings and Discussion

We played a sample of letters of the prepared dataset to the hearing-impaired individuals. The test showed a 100% understanding of the test data.

The results of playing words showed a 65% correct understanding of the played words.

The accuracy of the tool for understanding sentences was approximately 30%.

In the evaluation process, the person could recognize all shown letters successfully since they are clearly

Category	Entries
Kurdish Alphabets and Grammar	96
Greetings	13
Asking Questions	20
Colors	14
Time and Days	45
Opposites	134
Biology and Animals	105
Fruits and Vegetables	54
Food and drinks	87
Home	148
Family and Relatives	34
Human Body	48
Health and Hygiene	78
Clothes and Cosmetics	70
People with Special Needs	18
School	60
Verbs	193
Engineering and Mathematics	95
Science and Measuring Quantities	65
Art and Literature	25
Sound and Music	23
Sports and Games	56
History	27
Geography	50
Countries	78
City and Places	132
Transportation	63
Agriculture	46
Jobs	69
Economy	16
Finance and Currency	19
General Services	58
Institutes	17
Religion and Believes	52
Politics	41
Famous Personalities	11
Special Events	31
Other Words	124
Total	2315

Table 1: Kurdish Sign Language Dictionary Categories

shown in the dictionary. On the other hand, the signs for the words had a lower evaluation outcome. The person could not understand some of the words.

One reason for this was the usage of two different sign dictionaries in KRI. One of these dictionaries represents all signs based on the lexicon description, while the other (Ghazi Dizayee, 2000) uses vocal description for some of its entries. Our dataset was developed based on the latter. Both dictionaries are used interchangeably, but they provide different representations for specific signs depending on the context where they appear. This issue also affected sentence evaluation. Also, since we used a word by word translation, the hearing-impaired person was unable to understand

```
<sigml>
        <hns_sign gloss=<"ب ">
                <hamnosys_manual>
                        <hamflathand/>
                        <hamextfingeril/>
                        <hampalmdr/>
                        <hamtongue/>
                </hamnosys_manual>
 </hns_sign>
</sigml>
```

Figure 4: SiGML sample for letter "ب"

```
<sigml>
        <hns_sign gloss= <"زانکۆ">
                <hamnosys_manual>
                        <hamcee12/>
                        <hamthumbopenmod/>
                        <hamextfingerul/>
                        <hampalmd/>
                        <hamchin/>
                        <hamparbegin/>
                        <hamreplace/>
                        <hamextfingerul/>
                        <hampalmu/>
                        <hamparend/>
                        <hamlrbeside/>
                        <hamcheek/>
                </hamnosys_manual>
 </hns_sign>
</sigml>
```

Figure 5: SiGML sample for word "زانکۆ"

the meaning of a majority of the sentences as a whole. Therefore, the sentence evaluation achieved low accuracy, which is typical for the sign-supported systems.

7. Conclusion

We used HamNoSys to develop a sign dataset and its equivalent SiGML for Kurdish. We chose HamNoSys over SignWriting because of our plan to develop Kurdish Sign corpora in the future.

Our developed dataset includes approximately 560 entries consisting of the alphabet, numbers, and words. We also implemented a tool to translate Sorani texts into the Kurdish Sign language, which could be animated by an avatar.

We evaluated the tool by showing the animated output to hearing-impaired persons on the three aspects of understanding the sign gestures, namely letters, words, and sentences. The test showed a 100% understanding for the letters, a 65% for isolated words, and approxi-

mately 30% for sentences.
The main reasons for the low accuracy were the usage of more than one sign dictionary in the target community and the word-by-word translation of the input texts.
As future work, we are targeting the development of a language model based on the grammar of the Kurdish Sign language. Additionally, we aim to add more entries to the developed dataset. Furthermore, we would like to include other Kurdish dialects in the dataset.

8. Acknowledgements

We would like to appreciate the assistance of the Erbil Hearing-impairment Association for their assistance to access the Kurdish Sign dictionaries. Also, we acknowledge the help of Hiwa Center for Deaf and Mute in Erbil, particularly Ms. Mehan Fatah and Ms. Shno Aziz, for providing us with Kurdish Sign language resources and for their help in the evaluation process. Furthermore, we appreciate the constructive feedback we received from anonymous reviewers, which has helped us to improve the quality of the paper.

9. Bibliographical References

Ahmadi, S., Hassani, H., and McCrae, J. P. (2019). Towards Electronic Lexicography for the Kurdish language. In *Proceedings of the sixth biennial conference on electronic lexicography (eLex)*. eLex 2019.

Aouiti, N. (2013). Towards an automatic translation from arabic text to sign language. In *Fourth International Conference on Information and Communication Technology and Accessibility (ICTA)*, pages 1–4. IEEE.

Bouzid, Y. and Jemni, M. (2013a). An animated avatar to interpret SignWriting transcription. In *2013 International Conference on Electrical Engineering and Software Applications*, pages 1–5. IEEE.

Bouzid, Y. and Jemni, M. (2013b). An Avatar based approach for automatically interpreting a sign language notation. In *2013 IEEE 13th International Conference on Advanced Learning Technologies*, pages 92–94. IEEE.

Cooper, H., Holt, B., and Bowden, R. (2011). Sign language recognition. In *Visual Analysis of Humans*, pages 539–562. Springer.

Cormier, K., Fox, N., Woll, B., Zisserman, A., Camgöz, N. C., and Bowden, R. (2019). ExTOL: Automatic recognition of British Sign Language using the BSL Corpus. In *Proceedings of 6th Workshop on Sign Language Translation and Avatar Technology (SLTAT) 2019*. Universitat Hamburg.

da Rocha Costa, A. C. and Dimuro, G. P. (2001). SignWriting-Based Sign Language Processing. In *International Gesture Workshop*, pages 202–205. Springer.

Dhanjal, A. S. and Singh, W. (2019). Comparative Analysis of Sign Language Notation Systems for Indian Sign Language. In *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*, pages 1–6. IEEE.

Elliott, R., Glauert, J. R., Kennaway, J., Marshall, I., and Safar, E. (2008). Linguistic modelling and language-processing technologies for Avatar-based sign language presentation. *Universal Access in the Information Society*, 6(4):375–391.

Halawani, S. M. (2008). Arabic sign language translation system on mobile devices. *IJCSNS International Journal of Computer Science and Network Security*, 8(1):251–256.

Hanke, T. and Popescu, H. (2003). Intelligent sign editor. eSIGN project deliverable. *D2*, 3:2003.

Hanke, T. (2004). HamNoSys-representing sign language data in language resources and language processing contexts. In *LREC*, volume 4, pages 1–6.

Hashim, A. D. and Alizadeh, F. (2018). Kurdish Sign Language Recognition System. *UKH Journal of Science and Engineering*, 2(1):1–6.

Hassani, H. and Medjedovic, D. (2016). Automatic Kurdish dialects identification. *Computer Science & Information Technology*, 6(2):61–78.

Hassani, H. (2018). BLARK for multi-dialect languages: towards the Kurdish BLARK. *Language Resources and Evaluation*, 52(2):625–644.

Jepsen, J. B., De Clerck, G., Lutalo-Kiingi, S., and McGregor, W. B. (2015). *Sign languages of the world: A comparative handbook*. De Gruyter.

Kelly, D., Reilly Delannoy, J., Mc Donald, J., and Markham, C. (2009). A framework for continuous multimodal sign language recognition. In *Proceedings of the 2009 international conference on Multimodal interfaces*, pages 351–358.

Kipp, M., Heloir, A., and Nguyen, Q. (2011). Sign language avatars: Animation and comprehensibility. In *International Workshop on Intelligent Virtual Agents*, pages 113–126. Springer.

Liwicki, S. and Everingham, M. (2009). Automatic recognition of fingerspelled words in British Sign Language. In *2009 IEEE computer society conference on computer vision and pattern recognition workshops*, pages 50–57. IEEE.

McCarty, A. L. (2004). Notation systems for reading and writing sign language. *The Analysis of verbal behavior*, 20(1):129–134.

Mohammed, M. F. (2007). *The Sign Language for Deaf [In Arabic]*. Ministry of Labor and Social Affairs, Iraq.

Senghas, R. J. and Monaghan, L. (2002). Signs of their times: Deaf communities and the culture of language. *Annual Review of Anthropology*, 31(1):69–97.

Stiehl, D., Addams, L., Oliveira, L. S., Guimarães, C., and Britto, A. (2015). Towards a SignWriting recognition system. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 26–30. IEEE.

Sugandhi, P. K. and Kaur, S. (2018). Online Multilinguial Dictionary Using Hamburg Notaiton for Avatar-Based Indian Sign Language Generation System. *Int. J. Cogn. Lang. Sci.*, 12(8):1116–1122.

Wazalwar, S. S. and Shrawankar, U. (2017). Interpretation of sign language into English using NLP techniques. *Journal of Information and Optimization Sciences*, 38(6):895–910.

10. Language Resource References

Ghazi Dizayee, Adwiya. (2000). *Sign Dictionary for Hearing-Impairment [In Kurdish]*. The independent Human Rights Commission- Ministry of Labor and Social Affairs of Kurdistan Regional Government - Iraq- UNICEF- MEDS Organisation.

Nashat Salim, Nazanin and Hama Rashid, Jamil and Haji Omar, Salam. (2013). *Sign Dictionary for Hearing-Impairment [In Kurdish]*. The independent Human Rights Commission- Ministry of Labor and Social Affairs of Kurdistan Regional Government - Iraq.