An Arabic Sign Language Corpus for Instructional Language in School

Abdulaziz Almohimeed, Mike Wald, Robert Damper

School of Electronics and Computer Science, University of Southampton
SO17 1BJ, UK
{aia07r|mw|rid}@ecs.soton.ac.uk

Abstract

An annotated sign language corpus is essential for machine translation projects. For this reason, many sign language corpora have been developed. Unfortunately, none of these is based on Arabic Sign Language (ArSL). In this paper, we present the ArSL corpus we created that is based on school-level language instruction.

1. Introduction

In recent years, many efficient machine translation approaches, both statistical and example-based, have been proposed. These are corpus-based approaches. The accuracy of translation is directly correlated to the size and coverage of the corpus. The corpus is a collection of translation examples constructed from existing documents, such as books and newspapers. A written system for sign language (SL) comparable to that used for natural language has not been developed. Hence, no SL documents exist, which complicates the procedure of constructing an SL corpus. In countries such as the UK, Ireland, and Germany, a number of corpora have already been developed and used for machine translation (MT). Unfortunately, there is no existing Arabic Sign Language (ArSL) corpus for MT. Therefore, a new ArSL corpus for language instruction was created.

2. Recent Work

The following is a survey of recent work that has informed our project.

The Centre for Deaf Studies in the School of Linguistics, Speech, and Communication Sciences, Trinity College Dublin built an Irish sign language corpus (Leeson et al., 2006). This corpus, which contains children's stories, took approximately three years to build. There were 40 signers involved. The participants' ages ranged from 18 to 65 years, and they came from different regions in Ireland. The recorded videos are about 20 hours long. The videos were annotated using the EUDICO Linguistic Annotator (ELAN)[1]. The sign sentences were divided into different tiers that represent the Manual Features (MFs), referring to the hands, and Non-Manual Features (NMFs), referring to other parts of the body, such as the eyes, mouth, cheeks, etc., in gloss notation. In addition, an English translation was included for each sign sentence.

The European Cultural Heritage Online (ECHO) built a corpus of Swedish, British and Dutch SLs (Morrissey, 2008). It contains five children's stories signed in each SL. Approximately 500 signed sentences were collected in each language. ELAN was used to analyse the sentences.

Bungeroth et al. (2006) devised a German sign language corpus (DGS) for the weather report domain. They constructed their corpus by extracting the German subtitle text and DGS translation from a German daily weather news television channel called Phoenix Broadcasts. The signs were collected by extracting the lower right corner of the broadcast frame that shows the DGS interpreter. They used ELAN to analyse the DGS sentences. They separated these sentences into the following five tiers: gloss notation of the sign sentences, word classes (such as verb, noun, adjective, adverb, etc.), DGS sentence boundaries, German sentence translation, and German sentence boundaries. There were 2,468 sentences collected. This corpus was mainly designed for statistical machine translation and sign recognition.

3. Gloss Notation

A gloss notation is a textual representation of sign language. It is beneficial to use this notation method because it allows for storing and processing the signs, and a sign avatar can represent and animate the signs by passing the details of MFs and NMFs. Arabic letters will be used for the ArSL corpus annotation. The reason for this is that none of the signers assisting with the corpus building has the ability to write the gloss in English. Therefore, a new specification for writing the gloss notation in Arabic has been created.

NMFs will now be used to describe the use of gloss notation. Each NMF is represented as follows:

```
(NMF Part) -- "Action" -- Action
                Description
```

Example: (Mouth) – "جاء" – شد

where "جاء" means the signer is pronouncing the word "جاء" (i.e., "*jaa*") and شد represents the signer stretching the lips. Table 1 summarises all of the gloss notations used for the ArSL corpus.

An example of this is the textual representation of the sign sentence of the Arabic sentence "السرقه حرام":

```
(Mouth) شد "جاء"
(Head)
(Eyes) اغلاق
(Nose)
```

The empty tiers mean no action exists. These annotation tiers can be combined as

(Mouth) شد "جاء" (Eyes) اغلاق

[1] http://www.lat-mpi.eu/tools/elan/

Table 1: Summary of the Arabic gloss notation used in the ArSL corpus.

NMF	Action	Gloss
Eyes	Closing	“اغلاق”
	Opening	“فتح”
	Blinking	“ومض”
Nose	Wrinkling	“تجعد”
Mouth	Opening	“فتح”
	Closing	“اغلاق”
	Tongue out	“اخراج”
	Stre. lips	“شد”
	Sucking air	“شفط”
	Blowing air	“اخراج”
Shoulders	Forwards	“امام”
	Backwards	“خلف”
	Left	“يسار”
	Right	“يمين”
Cheeks	Puffing out	“مليء”
	Sucking in	“سحب”
Eyesbrows	Raising	“اعلي”
	Lowering	“اسفل”

where the empty features will not be taken into account.

4. Corpus Setup

4.1. Domain

The translation system still needs a suitable dataset. By restricting the corpus domain, the input sentences can be covered by the matching corpus sentences, which will increase the accuracy of the translation results. In addition, since each word can have more than one meaning, depending on the context, a restricted domain will help reduce this ambiguity. The constructed corpus domain was restricted to the instructional language that is used in schools for deaf students. It can be described as a one-directional instruction that communicates sentences from teachers to students. For this purpose, a corpus team was established that included three native ArSL signers and one expert interpreter.

4.2. Video Recording

To be sure that the translated sign sentences are fluent, clear, complete, and fully independent from the original Arabic sentences, the recording steps in Figure 1 were followed.
In Figure 1, sign sentences were produced after the interpreter showed the signers the meaning of the sentences using ArSL, without having them read the Arabic sentence. The reason is that after reading the Arabic sentence, the signers signed all of the original Arabic sentence details, even if they were not required. The signers also followed the order of the Arabic sentence. Then they signed it. After each sentence was recorded, the video was checked by the native signers to be sure that it was correct and would be

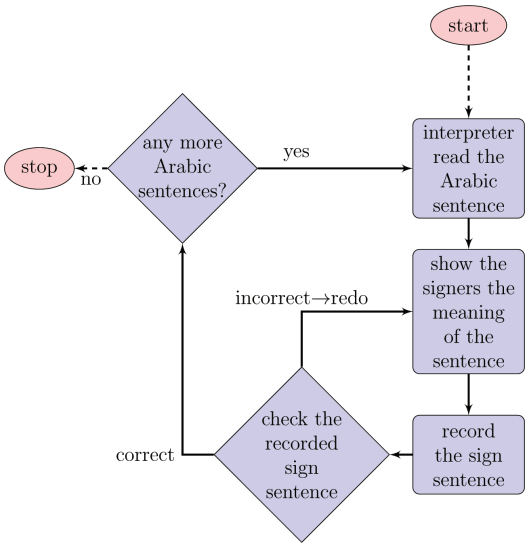

Figure 1: Steps in recording the signed sentences.

clear to deaf people in different age groups. If it was correct, they signed the next sentence; if not, the video was deleted and the sentence was recorded again. In the end, 213 ArSL sentences were recorded using a Sony DSC-W120 digital camera and were stored in MPEG format. The size of the recorded video frame is 640 × 480 pixels.

4.3. Video Corpus

After recording the sign sentences, the videos were annotated using the ELAN annotation tool. As shown in Figure 2, Arabic translation was added. Then, boundaries for each sign in the recorded video were clearly marked, and extra information was added. This information contained both MFs and NMFs. NMFs were described using the gloss notation discussed above. After isolating and adding the MFs and NMFs for all of the signs in the ArSL sentences, the annotated ArSL data were saved in EAF XML format.

4.4. Bilingual Corpus and ArSL Sign Dictionary

After the annotated ArSL data were saved in EAF XML format, the next phase was to build a bilingual corpus of ArSL and Arabic text delivered from the EAF and MPEG files. This procedure is essential for ArSL translation. The first step in constructing the bilingual corpus, is parsing the EAF XML files and extracting the MFs and NMFs for each sign, as shown in Figure 3.
Considering the information extracted for each feature (see bottom of Figure 3), the feature name field determines which part of the body is being used (this may be the right hand, left hand, mouth, etc.); the text shows the gloss notation for the particular body part. The EAF file name and Video fields identify the EAF and Video locations for each part. The start and finish time determines the exact location of the feature in the source video, which will be used later in constructing the signs-to-Arabic dictionary to extract the sign video clip from the source video. After the completion of this step, 1,897 features had been extracted. The next step in constructing the bilingual corpus is producing the sign dictionary using the extracted MFs

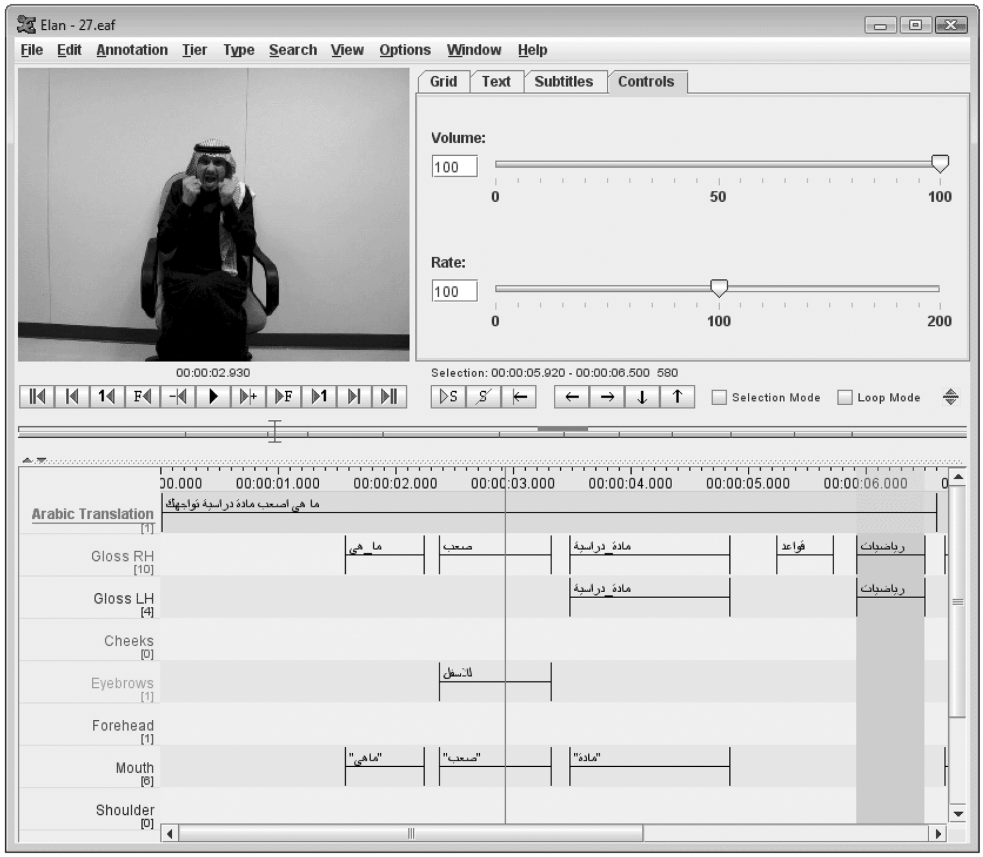

Figure 2: Using the ELAN Linguistic Annotator to annotate sign sentences.

```
<TIER TIER_ID="Arabic Translation" LINGUISTIC_TYPE_REF="default-lt" DEFAULT_LOCALE="en">
    <ANNOTATION>
        <ALIGNABLE_ANNOTATION ANNOTATION_ID="a9" TIME_SLOT_REF1="ts1" TIME_SLOT_REF2="ts2
      <ANNOTATION_VALUE>لا تضرب من هو اصغر منك</ANNOTATION_VALUE>
    </ALIGNABLE_ANNOTATION>
  </ANNOTATION>
</TIER>
<TIER TIER_ID="Gloss RH" LINGUISTIC_TYPE_REF="default-lt" DEFAULT_LOCALE="en">
  <ANNOTATION>
    <ALIGNABLE_ANNOTATION ANNOTATION_ID="a10" TIME_SLOT_REF1="ts3" TIME_SLOT_REF2="ts6">
      <ANNOTATION_VALUE>لا</ANNOTATION_VALUE>
    </ALIGNABLE_ANNOTATION>
  </ANNOTATION>
  <ANNOTATION>
    <ALIGNABLE_ANNOTATION ANNOTATION_ID="a11" TIME_SLOT_REF1="ts9" TIME_SLOT_REF2="ts11">
      <ANNOTATION_VALUE>تضرب</ANNOTATION_VALUE>
    </ALIGNABLE_ANNOTATION>
  </ANNOTATION>
  <ANNOTATION>
    <ALIGNABLE_ANNOTATION ANNOTATION_ID="a12" TIME_SLOT_REF1="ts13" TIME_SLOT_REF2="ts24">
      <ANNOTATION_VALUE>اصغر_منك</ANNOTATION_VALUE>
    </ALIGNABLE_ANNOTATION>
  </ANNOTATION>
</TIER>
<TIER TIER_ID="Gloss LH" LINGUISTIC_TYPE_REF="default-lt" DEFAULT_LOCALE="en"/>
<TIER TIER_ID="Cheeks" LINGUISTIC_TYPE_REF="default-lt" DEFAULT_LOCALE="en"/>
<TIER TIER_ID="Eyebrows" LINGUISTIC_TYPE_REF="default-lt" DEFAULT_LOCALE="en">
  <ANNOTATION>
    <ALIGNABLE_ANNOTATION ANNOTATION_ID="a13" TIME_SLOT_REF1="ts4" TIME_SLOT_REF2="ts7">
      <ANNOTATION_VALUE>"تقوس للاسفل"</ANNOTATION_VALUE>
    </ALIGNABLE_ANNOTATION>
  </ANNOTATION>
</TIER>
<TIER TIER_ID="Forehead" LINGUISTIC_TYPE_REF="default-lt" DEFAULT_LOCALE="en"/>
<TIER TIER_ID="Mouth" LINGUISTIC_TYPE_REF="default-lt" DEFAULT_LOCALE="en">
  <ANNOTATION>
    <ALIGNABLE_ANNOTATION ANNOTATION_ID="a14" TIME_SLOT_REF1="ts5" TIME_SLOT_REF2="ts8">
      <ANNOTATION_VALUE>"لا"</ANNOTATION_VALUE>
    </ALIGNABLE_ANNOTATION>
  </ANNOTATION>
```

Features | MF & NMF Table

ID	eafFileName	VideoFileNam	Part	aID	Stim	Ftim	txt
2175	D:\Corpus\EAF\9.	MOV00011.MPG	Arabic Translation	a9	0	7524	لا تضرب من هو اصغر منك
2176	D:\Corpus\EAF\9.	MOV00011.MPG	Gloss RH	a10	1140	2320	لا
2177	D:\Corpus\EAF\9.	MOV00011.MPG	Gloss RH	a11	2490	3280	تضرب
2178	D:\Corpus\EAF\9.	MOV00011.MPG	Gloss RH	a12	4200	6100	اصغر_منك
2179	D:\Corpus\EAF\9.	MOV00011.MPG	Eyebrows	a13	1140	2320	"تقوس للأسفل"
2180	D:\Corpus\EAF\9.	MOV00011.MPG	Mouth	a14	1140	2320	"لا"
2181	D:\Corpus\EAF\9.	MOV00011.MPG	Mouth	a15	2490	3280	"تضرب"
2182	D:\Corpus\EAF\9.	MOV00011.MPG	Mouth	a16	4200	6100	"اصغر منك"
2183	D:\Corpus\EAF\9.	MOV00011.MPG	Shoulder	a17	4390	5210	انحناء

Figure 3: Parsing EAF XML.

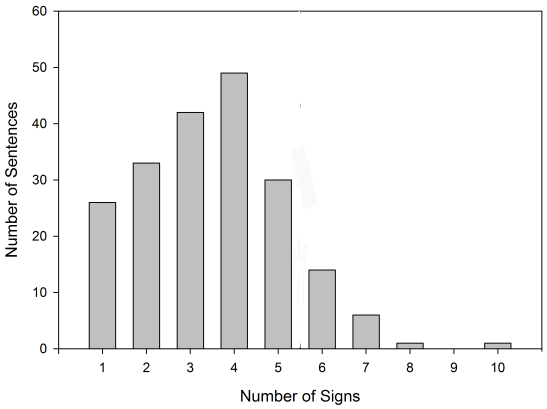

Figure 4: Distribution of collected sentences according to the number of signs that they contain.

and NMFs. All features that occur in the same period of time and have the same video source are considered to belong to same sign, and are collected together with that sign. Arabic sentence translation is then used to produce sign sentences in Arabic (bilingual corpus). The signs in this sentence will be linked to the corpus in the correct order using the video name.

The next step is extracting the video clips from the source video files using the start and finish times. After extracting each clip, the clip location will be appended to the sign table.

The last step is adding tags to represent the syntactic and morphological information for each sentence. The following is an example:

Arabic Sentence: لا تضرب من هو اصغر منک

After adding tags: <particle> لا

<verb present> تضرب

<preposition> من

<personal pronoun> هو

<adjective> اصغر

<preposition> من+

<personal pronoun> ک

In the end, there were 710 signs in the dictionary. There were 203 signed sentences in the bilingual corpus. The dsitribution of sentences according to the number of signs that they contain is shown in Figure 4.

5. Conclusion

We have presented an ArSL corpus for school-level language instruction. The corpus contains two main parts. The first part is the annotated video data that contains isolated signs with detailed information that includes MFs and NMFs. It also contains the Arabic translation script. The second part is the bilingual corpus that is delivered from the annotated video. A translation system can be used with a bilingual corpus. The ArSL corpus is now publicly available from `www.ArSL.org` and is suitable for ArSL recognition and translation systems. We are currently using the corpus to conduct translation experiments with Arabic text. We also plan to extend the number of examples to cover a larger domain.

Acknowledgments

This corpus could not have been constructed without the hard work of our expert signers: Ahmed Alzaharani, Kalwfah Alshehri, Abdulhadi Alharbi and Ali Alholafi.

6. References

J. Bungeroth, D. Stein, P. Dreuw, M. Zahedi, and H. Ney. 2006. A german sign language corpus of the domain weather report. In *Fifth International Conference on Language Resources andEvaluation*, pages 2000–2003, Genoa, Italy.

L. Leeson, J. Saeed, C. Leonard, and A. Macduff andD. Byrne-Dunne. 2006. Moving heads and moving hands: Developing a digital corpus ofIrish sign language. 'the signs of ireland' corpus developmentproject. In *Information Technology and Technoloy Conference 2006*, pages 33–43, Carlow, Ireland.

S. Morrissey. 2008. *Data-Driven Machine Translation for Sign Languages*. Ph.D. thesis, School of Computer Science, Dublin City University, Ireland.