Translation System from Arabic Text to Arabic Sign Language

Nadia Aouiti[1], Mohamed Jemni[2]
Research Laboratory LaTICE, University of Tunis 5, Avenue Taha Hussein, B. P. : 56, Bab Menara, 1008 Tunis, Tunisia.
nadia.aouiti@gmail.com[1], mohamed.jemni@fst.rnu.tn[2]

Abstract - *This research paper presents our ongoing project aiming at translating in real time an Arabic text to Arabic Sign Language (ArSL). This project is a part of a Web application based on the technology of the avatar (animation in the virtual world). The input of the system is a text in natural language. The output is a real-time and online interpretation in sign language. Our work focuses on the Arabic language as the text in the input, which needs many treatments due to the particularity of this language. Our solution starts from the linguistic treatment of the Arabic sentence, passing through the definition and the generation of Arabic Annotation Gloss system and coming finally to the generation of an animated sentence using the avatar technology.*

Keywords- *Arabic Language, Arabic Sign Language, translation, Arabic Gloss.*

1. INTRODUCTION

Sign Language (SL) is a natural language for deaf people. It is a body language used for communication and serves as the same function than spoken language. In the last century, SL gains increased attention by scientists in the fields of language and computer sciences. However, a few of them were interested in ArSL, due to the difficulty of the Arabic language, especially with the complexity of morphological and grammatical rules. Unfortunately, 25% of the community of deaf person around the world is in the Arabic world [1]. This very important number has directed us to try to develop an Arabic system for Arabic deaf people, taking as input, Arabic text, and giving, as a result, an animated sentence using the avatar. In this context, our paper presents a description of our proposed translation system named ArabSTS and based on Arabic Gloss annotation system for ArSL. The rest of this paper is structured as follows: In the second section, we present an overview about a various works related to our project for different Sign Languages. The third section focus on the transcription systems used for the translation of the text into SL. The fourth section is devoted to presenting the different problems related to the language in input which is Arabic language and its particularity. The fifth section presents the **ArabSTS (Arab S**ign **L**anguage **T**ranslating **S**ystem**)** with a description of each phase composing our project.

2. RELATED WORK: EXISTING SYSTEMS

In the last years, Sign Language gained increased attention by researchers. Many works have been developed for many languages. In fact, we can divide them into two groups referring to the language in input: Machine Translation for Arabic language and Machine Translation for non-Arabic language.
Several projects have been done taking as input non-Arabic language such as English, French, Japanese, British, etc. In what follows, we present some example of translating systems.

2.1 TESSA

It is a Speech-To-British Sign Language (BSL) translation system that aims to provide a communication aid between a deaf person and a Post Office clerk[2]. The system uses a formulaic grammar approach where a set of pre-defined phrases are stored and the translation is done by using a phrase lookup table. However, the use of a small set of sentences as templates makes TESSA a very domain specific system. It assumes a very restricted discourse between the participants[3].

2.2 TEAM Project

This project was proposed by Zhao, K. Kipper, W. Schuler, C. Vogler, N. Badler, and M. Palmer [4]. It's a Text-To-ASL translation system where the STAG (Synchronous Tree Adjoining Grammar) formalism is used to represent source text into the ASL syntactic structure. The system maintains a bilingual lexicon to identify the valid word-sign pair. The output of the linguistic module was a written ASL gloss notation. The manual and non-manual information, including the morphological variation, are embedded within the ASL gloss notation. The output of the synthesis module uses animated human models (Avatar)[5].

2.3 ViSiCAST Translator

Marshall and al at the University of East Anglia implemented a system for translating from English text into British Sign Language (BSL)[6]. Their approach used the CMU Link Parser to analyze the input English text. And they used Prolog declarative clause grammar rules to convert this linkage output into a Discourse Representation Structure. During the generation half of the translation process, Head-Driven Phrase Structure rules are used to produce a symbolic SL representation script. This script is in the system's proprietary, a symbolic coding scheme for the movements required to perform a natural Sign Language [7].

In addition, (Morrissey and Way, 2005, 2006) and (Morrissey et al, 2007) proposed a system using example-based methodologies as part of a data-driven framework. (Stein et al, 2006) has proposed a statistical MT system which uses Hidden Markov Model and IBM models for training the data. (Chiu et al, 2007) also, present a Statistical approach for their work with Chinese and Taiwanese Sign Language [3].

All the projects cited above was a part of the non-Arabic machine translation systems. For Arabic language the existing projects are rare, for instance, we have:

2.4 A. Almasoud and H. Al-Khalifa

Their system aims to translate an Arabic text into Arabic sign language. This translation solution [8] is based on the translation rules of the Arabic sign language and the domain ontology to produce the notation "SignWriting" (which is an annotation system for the sign language we go to Detailed in the next chapter). The "SignWriting" will then be used as the final output of the system or as an intermediate level for the future animation of the avatar.

An additional level of semantic processing is added by this system using the Semantic Web technology. However, this proposed semantic translation system is limited to the jurisprudence of prayer, as it is a small area with the limited vocabulary that is necessary for Arab deaf Muslims.

2.5 Mohandes

It is a system [9] developed to translate Arabic text into Arabic SL. This system is one stage in the process of developing a system to translate Arabic speech to Arabic sign language.

The system has a database, to store a dictionary containing the words in Arabic with their corresponding signs. Each sign is recorded by a sign language expert in a video sequence. Images and video files representing words in the Arabic dictionary are stored physically on the server in a special location. When the server receives a request for a particular sign, it will always search for the video in that location. The symbols are then represented in the dictionary through the name of the corresponding file in the video sequence. If the user enters an available word in the database, the recorded sequence will be displayed. Otherwise, if the word does not exist in the dictionary, the spelling with the fingers will be done (Finger Spelling).

2.6 Tawassol

"Tawasol" is a free educational application, carried out within the Tawassol center. This application has been developed in the field of assistive technology and education to support people with hearing impairment.

It is another Arabic system [10][11] for translating Arabic text into Arabic SL. The system is used as an educational tool. It contains a translator, a dictionary of Arabic words in a set of categories, in addition to a finger spelling editor. The system uses Vcommunicator Gesture Builder 2.0 with Sign Smith Studio program to generate the animation output.

Tawasol app contains three main functions:
- Dictionary: it contains a different theme in which covers an important number of words. The dictionary function offers for the user the possibility to find a word, to repeat the word, to have sound control.
- Tutorial: it's a quiz tutorial in which the user have to match the word with the right video of the sign. The user has the ability to modify the level of difficulty of the quiz, the number of the word, etc.
- Finger-spelling: the system takes the word entered by the user and displays the sign language symbol for each letter.

2.7 Almohimed, Wald and Damper

This translating system[12] adopts a chunk-based Example on Based Machine Translation (EBMT) system which produces output sign sentences by comparing the Arabic text input to matching text fragments, or "chunks". The corpus of this system has 203 signed sentences with content restricted to the domain of instructional language, typically used in deaf education.

The system has two phases. Phase1 is run only once; it pre-compiles the chunks and their associated signs. Phase 2 is the translation system, that converts Arabic input into ArSL output using Arabic Gloss annotation system only for facial expression (Mouth, Head, Eyes, and Nose)[5].

3. THE ANNOTATION SYSTEMS: STATE OF THE ART

Due to the complexity of the translation process of SL, to translate a natural language (text or sound) to SL, we have to use one of the transcription systems. Three main outputs for the machine translation to SL exist such as:
- The Visual: the Avatar or the video sequences;
- The Symbols: the notation system like: Stockoe notation system, SignWriting, HamNoSys, etc;
- The Gloss: the textual representation.
In this section, we will present some of the existing transcription system.

3.1 Stockoe notation System

This system is invented in 1960 by William Stockoe. He was the first linguistic which proposes three parameters to describe the sign such as:
- The location of the sign identified by the symbol "tab" (Tabula);
- The hand shape identified by the symbol "dez" (designation);
- The movement of the sign identified by the symbol "sig" (Signation).

3.2 Hambourg notation System

This notation system was created in 1985 by a research team in Hamburg Center for German Sign Language. HamNosys is a national system which serves to transcribe signs for sign language research.
To describe a sign in HamNoSys, we have to respect five parameters:
- Hand Shape (Fist, Flat Hand, Separated Fingers, Thumb Combination);
- Hand Orientation (Wrist, Extended Finger and Palm Orientations);
- Location (refers to the location on the body);
- Movement (Straight, Curved, Wavy, zigzag, circular and spiral);
- NonManuals Signs.

3.3 Sutton SignWriting notation System

This system was created in 1974 by a research team in Copenhagen University. The SignWriting (SW) international alphabet named ISWA contains 630 basic symbols divided into seven categories which are: Configuration, Movement, Figure and head, Body, Dynamics and rhythm, Punctuation and Advanced subdivision.

The symbols used in SW allow describing the shape of the configurations, the trajectory of the movements, the facial expressions and the movements of the head or the torso.

3.4 Gloss notation System

The Gloss notation system, which is used in the proposed system, is a textual representation accompanied with some of the symbols which allow identifying a sign. It covers the different parameters of the sign with a simplified representation to avoid the complexity of understanding.

Each Gloss is represented by one word in the capital letter. Each Gloss can be accompanied with transcription symbols to express different categories of linguistic and sign information as following:
- A topic of the sentence;
- Type of the sentence;
- Location of the sign;
- Number of repetition of the sign;
- Hand shape;
- Compound or contraction sign;
- Emphasized sign;
- Fingerspelled Sign;
- Arc motion signs;
- Facial Expression.

All these information are indispensable for the translation step. More we can provide linguistic, morphologic and Sign information, more the Gloss notation will be efficient and util.

3.5 Discussion

Figure1. Examples of notation systems for SL.

All these notation system cited below, are developed specially for the transcription of the sign in Sign Language. Each system has its own characteristic which differed from the other systems. The effectiveness of each of them depends mainly on several factors such as The input language, the domain of use of the notation system, the facility of use, etc.
In our proposed system, we have used the Gloss annotation system. The Gloss annotation system is a textual representation which is used as an intermediate level between written text and SL for the machine translation. This transcription system is considered as the most simplified one. All the other annotation systems suffer from the complexity and the difficulty of use.

4. ARABIC LANGUAGE AND PARTICULARITIES

Arabic is a Semitic language, written from right to left and composed of 28 letters. It is classified as one of the difficult languages. The difficulty of this language comes from many reasons related to its characteristics.

The particularity of this language comes mostly from the presence of diacritical marks that are added above or below the letter. These marks are required to read and indicate the right meaning of the text. They help to differentiate between words with the same representation or form. However, the diacritical marks are only used for sacred and teaching texts. The absence of the vowel in Arabic text causes also, an ambiguity with two levels:
- The meaning of the word: for example, the letter "ى" can be written with diacritic marks, and in this case, it means the letter "ي" (ya), or without diacritic marks and that means the letter "ى" (a);
- The difficulty to identify the right position of the word in the sentence [13]: The absence of the vowel in the Arabic sentence causes a big problem to determinate the right function of the word. To more explain this problem we have occurred to present an example which is illustrated in the figure below.

The lack of capital letters presents also, a problem for Arabic to identify proper nouns, abbreviations, acronyms, and the adjectives. This problem causes a big ambiguity to differentiate between adjective, proper noun, and noun. Taking the example of the word " سعيد " which can easily generate an ambiguity to determine its main function. It can mean on one hand, the adjective "happy" and on the other hand the proper noun "Saiid". Therefore, the absence of capital letters in Arabic language causes the non determination of the appropriate function of the word in the sentence. Arabic language is a complex language that needs a lot of treatments to analyze and identify each word which composes the sentence. The particularity of this language affects the automatic processing and produces a strong ambiguity. The following example illustrates the problems cited above:
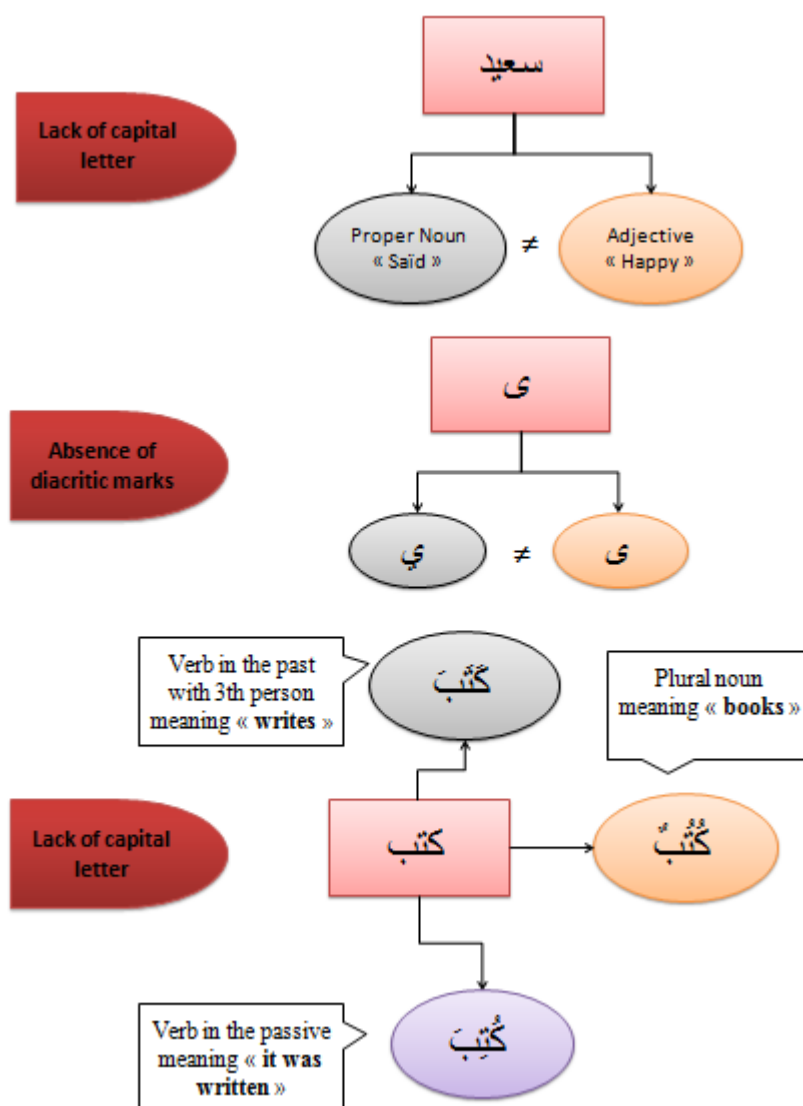
Figure 2. Example of problems due to the absence of capital letter and diacritic marks.

5. OUR SOLUTION: "ArabSTS"

Our system, called **A**rab**STS** (**A**rabic **S**ign **L**anguage **T**ranslation **S**ystem), aims to translate Arabic text to ArSL. This translation contains many modules. Each one serves as an input for the other, starting from the automatic treatment of Arabic text attending an animated avatar ready to translate the sentence.

To attempt this target, our proposed system includes many steps that will be described in details in this section. To more explain our project, we present in the following the global architecture of our system containing the main phases:
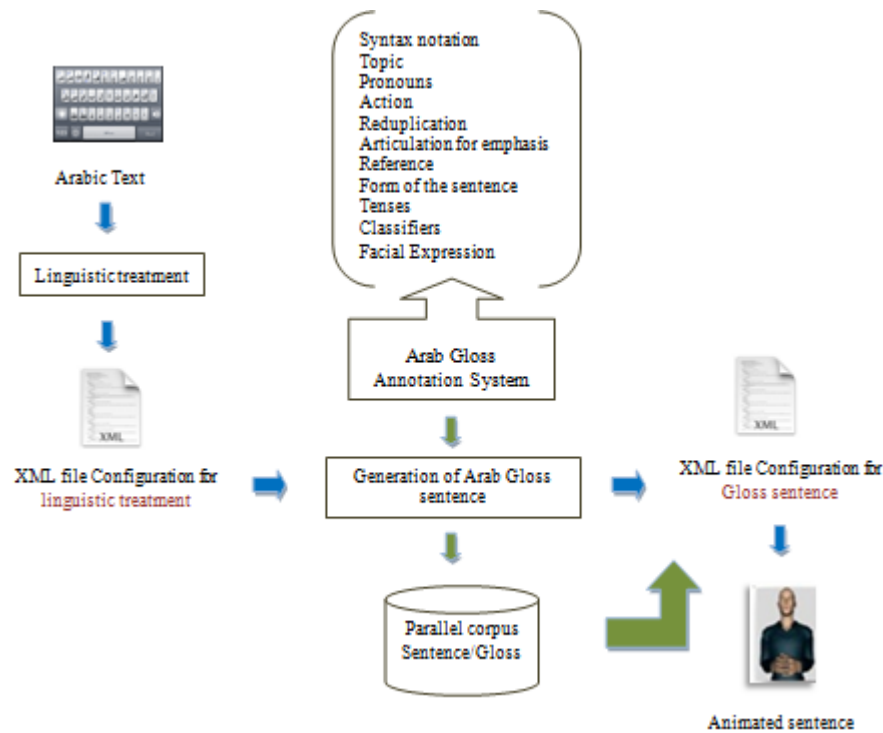
Figure 3. Global Architecture of our project.

Our project takes as input an Arabic text, This text is then a subject of an important linguistic treatment using the MADAMIRA Analyzer system. The result of this step will be an XML file configuration in which we have the different parts of the sentence with function and necessary information (subject, noun, punctuation, number of segment, pronouns, …).

On the other hand, we define the Arab Gloss annotation system as an intermediate text representation. Arab Gloss defines an important number of annotation rules to cover the maximum of information relevant for the translation step.

The definition of the Arab Gloss transcription system was based on extensive studies of the existing annotation system, the Gloss annotation system for the other language and the different rules of the Arabic Sign Language existing.

Referring to this transcription system, we build the parallel corpus composed by the Arabic sentence with its Arabic Gloss representation. This step is used necessarily to minimize the time needs to translate the sentence.

If the sentence in input was found, we will generate the XML file configuration for Gloss sentence to animate the avatar, else we will use the rules defined in the Gloss annotation system to generate the Gloss sentence. This database is after alimented with all the sentences treated by this system. In the following, a detailed description of each phase of our project is presented.

5.1 Linguistic treatment of Arabic text

The process of language treatment is divided into several stages. First, the original text undergoes a morpho-syntactic analysis using a well-specified lexicon, grammatical and morphological rules. The result of this step is a syntactic representation of level 1 that will be analyzed semantically[13] [11] to achieve a representation of level 2.

The process of the automatic treatment of Arabic is very complex. Different tools are developed for this target and that differ by their rate of precision to identify each word in the sentence. However, until now, there are no efficient tools that can give a rate of precision equal to the human result in analyzing Arabic language. A limited number of systems for morphological analysis and disambiguation of Arabic exist, and few of them are used for free. For our project, we tried different morphological systems in order to choose the most adequate for our work. For

instance, we have firstly tested Stanford Parser [14] [4], available for free, which ensures the segmentation of a sentence and returns the set of segmented words with their grammatical category. This information is important to reduce the error of translation of sentences. However, this system has a problem in the treatment of some case of Arabic language. For example, the word: «كتابه» (his book), which is written in one word and composed by the noun (book) and the attached pronoun (ه). Using Stanford Parser, we cannot solve this problem and parse this word to define the right function of each component of the word (noun + attached pronoun), and this is very important to give the appropriate translation of the sentence.

Also, we have tested MADAMIRA [15] parser system. The Input text (either MSA or EGY) enters the Preprocessor, which cleans the text and converts it to the Buckwalter representation used within MADAMIRA. The text is then passed to the Morphological Analysis component, which develops a list of all possible analyses for each word. The text and analyses are then submitted to a Feature Modeling component, which applies SVM and language models to derive predictions for the word's morphological features. SVMs are used for closed-class features, while language models predict open class features such as lemma and diacritic forms. A score is given then to each analysis. The top-scoring analysis passed then to the tokenization to generate customized tokenization for the word according to the schemes requested by the user.

When all the requested components have finished, the results are returned to the user. Users can request specifically what information they would like to receive; in addition to tokenization, base phrase chunks and named entities, the diacritic forms, lemmas, glosses, morphological features, parts-of-speech, and stems are all directly provided by the chosen analysis. The output of this system is as follows:

Figure4. Output of the sentence « أكل الولد التفاحة » (the boy eat the apple) using MADAMIRA parser system.

The output of this system gives a number of information which is very useful for our work. Taking the example above, MADAMIRA gives the different functions of each word composing the sentence such as gender (masculine or feminine), number (plural or singular). If it is a verb, as we see in the example at the top, we have also the person with whom it is conjugated. In addition to that, concerning the problem of attachment between the noun and the possessive determiners (تفاحته, his apple), MADAMIRA parses the word and divides it into two parts giving the right function of each one. The XML file contains a segmentation of the sentence into words, then word information part containing all the details about each segment (function, number, gender, etc).
This output is after used to generate an XML file containing a list of information related to each word composed the sentence to be used as an input of the XML file configuration for Gloss sentence.

5.2 Definition of Arabic Gloss Annotation System: Our solution

As we have cited earlier, only a few works have been devoted to Arabic language[16], especially with the complexity of morphological and grammatical Arabic rules. In this context, our work focused in the definition of an intermediate representation between Arabic text and ArSL covering the maximum of rules and information necessary for the translation step. This annotation system is a text representation that covers not only morphological and grammatical rules, but also hand-shape, sign location, facial expression and movement[17].

By analogy to ASL Gloss annotation system, we proposed an Arabic Gloss annotation system referring to all guidelines outlined in the agreement of Liddell [18] and based on the Arabic and Quatarian Sign Language rules[19].

Our Arab Gloss proposed system [4] covers many linguistic information, indicators for facial expressions, locations of the sign, positions, etc. In this section, we will present an overview of this transcription system with some examples.

5.2.1 Syntax and morphology rules for ArSL Gloss

Each Gloss is represented by an Arabic word. One Arabic word identifies one Arabic sign. In Arab Gloss, for some case, even the form of the word can be changed. We can add a symbol in the beginning, the middle or the end of the word. Each symbol is used to give a meaning to the word. For example, the Hyphens "-" added between the letters of the word is used to spell the word. However, if it is added between two words, this means that it indicates two ArSL signs have been combined into a single one. The symbol Slash "/" also used if a sign has two different words in Arabic. Hash "#" before the word is used to be fingerspelled, and in general, it is used to spell the names. "!" exclamation point, is added before and after the word to indicate a stressed articulation for emphasis and therefore, the sign is a bit faster, stronger, or more exaggerated than normal. For example the gloss ‎!سهل‎ (easy) is a transcription notation for the word «‎سهل جدا‎» (very easy) with a stressed articulation that occurs simultaneously with the manual signature. An overview of these symbols with the meaning of each one is presented in the following figure:

Figure 5.Example of Syntax Arab Gloss annotation

In addition to the syntax Gloss annotation, we define the different rules to indicate the: topic of the sentence, reduplication of the sign, tenses, pronouns, action, the form of the sentence, etc. To better present the different rules composing our annotation system, we present, in the following, an example that covers a number of Gloss used to transcribe an Arabic sentence to Arab Gloss:
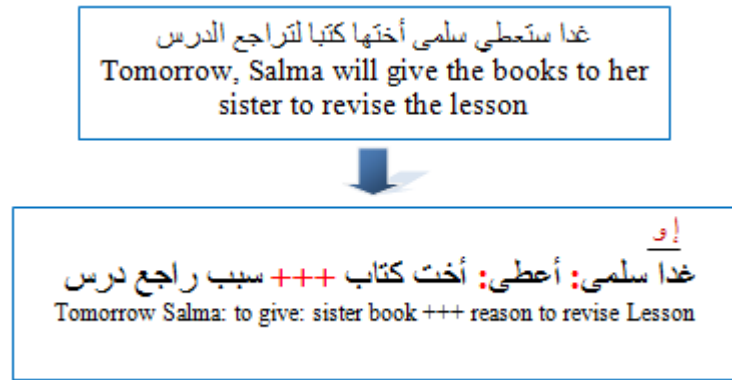
65

Figure 6.Transcription of the sentence "غدا ستعطي سلمى أختها كتبا لتراجع الدرس" in Arab Gloss annotation

The example presented above covers many transcription rules. In this transcribed sentence, we note that all the verbs were in the infinitive form (أعطى, راجع). In ArSL, we do not conjugate verbs. So, translating sentence contains only verbs in the infinitive form. Also, the presence of the symbol "**+**" may be repeated many times successively. The number of "+" indicates the number of repetitions of the sign. In fact, two cases are figured for glossing and this depends of the number of repetitions (two times and three times). In general, two plus symbols are used for the noun in the plural form which presents two things (Two books, two boys, etc.). Whereas, if the symbol plus is repeated three times, it indicates that the noun is in plural form and for more than two things, so the sign will be also repeated three times. In the other hand, as we defined the number of the repetition of the sign, the position of the sign is also very important. Therefore, for some verbs the position depends on the verb, subject and object. "i:GLOSS:j" this notation is used to indicate that the sign is directed from the first person to the second person. For example, the verb أعطى(give) would begin its articulation in the point in space associated with the subject referent (call that location أ(i)) and would finish its articulation in the location associated with the object referent (call that location ب(j)) as follows: أعطى: أ[5].ب .

The line above the word indicated by the symbol "إو"meaning that this word is a tense indicator. It is placed at the beginning of the sentence to set up the tense. The other possibility to transcribe the tense in ArSL Gloss, is the adding of some words such as: "إنتهى"(is finished), "بعد"(after). These words are used when there is no indicator of time in the sentence.

The annotation of the line above the sentence is used also to indicate the form of the sentence. In fact, we have four types which are:

- Interrogative form, marked with the letters "س-م ز س"for the question sentence to answer about place, time or reason. Also, we have the letters "س-نعم-لا"this annotation is used for the Yes or No question.
- Conditional form, marked with the word **"شرط"**.
- Negative form: In glossing negative form is indicated by the symbol "سلب". It is a headshake accompanied by a facial expression meaning the negation.
- Rhetorical question: It is a form of a question that is asked in order to make a point rather than to elicit an answer. For glossing rhetorical question is marked by the symbol " س-أنك".

In addition to all these rules defined below for ArSL Gloss annotation system, pronouns are also redefined. Therefore, pronouns are divided in three categories Addressee pronoun (المخاطب), Speaker pronoun (المتكلم) and Absent pronoun (الغائب).In each category, we have the first person (singular masculine/feminine form) and others (two or more than two people in the feminine or masculine form). Taking the example of the sentence "أنت ذكي" (you are smart), the transcription of this sentence in ArSL Gloss will be: ذكي **1-مخ-ض** . In this example, we have used the addressee pronoun for the first person category.

5.2.2 Classifiers in ArSL Gloss:

In this system, we transcribed not only syntax and morphological function of ArSL but also, movement, form and size of the objects. It is what we call Classifiers in ASL [20]. Each classifier, presents an additional information about the subject (noun or object). It is a set of information to describe the form of the object, the object's moving, the relation of the object with another object or person, etc. In fact, in ASL many types of classifiers exist such as: Semantic classifiers, Body classifiers, Hand shape classifiers, etc.

At this level, we have defined the hand shape classifier used for ArSL. By analogy to ASL, Classifiers are represented with the word "صنف". Additional Arabic letters or numbers are added to each one to give descriptive information about the subject or the predicate (verb). For example, the classifier "صنف ن" as the picture below, can mean a small and around thing, the width or the height of one object, etc. Such as: button, disc, moon, small bottle, etc.

Figure7. Hand shape classifiers used for ArSL

Each type of classifier is used to describe a group of things having a common hand-shape. As shown in the figure above, we have 10 proposed hand-shape classifiers. The main rule of this category of classifiers is to describe the size, the form, the movement of the object, thus to more clarify the sentence as follows: [4]

- Classifier «ن- مغير» صنف:
  - Long cylindrical objects: hose, etc.
  - Viewing devices: goggles, telescope, camera, etc.
  - Short cylindrical objects: cup, glass, bottle, etc.
- Classifier «ن-مسطحة» صنف:
  - Short object: small stack of papers, etc.
  - Flat object with squared edges: ruler, paper, etc.
  - Thin object
- Classifier «صنف-خ»:
  - Describing a flat object with position relative to the signer: roof, tunnel, street, etc.
- Classifier «صنف-ح»:
  - Curved objects: bowl, sink, basin or a digging device
- Classifier «صنف-س»:
  - Smooth, flat surfaces: road, window, wall, door, land, mountain, table, etc.
  - Smooth, flat surfaces: a long stretch of desert or sky, etc.
  - Flat mobile surfaces: valley
  - To give some description: person's height (tall or short), light, etc.
  - The position of an object or a person: Distant, near, before or behind, above, next to, outside, etc.
- Classifier «صنف-ه»:
  - Solid, spherical objects.
  - Handles conveyance: Motocycle, bicycle, etc.

5.3 Interface of our system

The following image represents our welcome interface for our system aiming to translate Arabic text to Arabic Sign Language using Arabic Gloss.
Our interface gives the hand to enter a sentence written in Arabic text. Then, ArabSTS analyzes this sentence using the analyzer MADAMIRA cited in the previous section.

The output of this analyzer is used to generate the Arabic glosses which gives as the result an XML file containing the necessary information used for the animation of the avatar such as (reordering the words composing the sentence, the word will be signed or not, stressed sign, position of the sign, repeated sign, finger spelled sign, form of the sentence, etc).

The following interface takes as an example the sentence " كم الساعة ؟ " "**What time is it?**", ArabSTS gives as the result an XML File for Gloss sentence which is transcript and ready to be translated to Sign Language.

Figure 8. Home interface for ArabSTS

5.4 Implementation of Arab Gloss Annotation System

The system takes as an input an Arabic sentence, this sentence will be parsed and the system determinates the several parameters depending to the type of the sentence (interrogative form, passive form, starting with a noun or a verb, etc).

For each type of sentence, we will ask then for the function of each word composing the sentence. After an important treatment using the linguistic analyzer and extracting all the necessary information, we have in the result an XML file configuration containing two parts:
- Part1: for each word, we give the (id, genre, num, function, indication). The indication field is the result of the previous treatment. It may have the value (0,No-sign, #, / , !, etc) which means respectively (No indication, No-sign to translate, word finger spelled, the same sign for two different words, stressed marks, etc).
- Part2: this part gives the final form of the sentence ready to be translated: the order of the word to sign, delete some words, two words combined in one sign, etc, using the indications given in the previous part.

This XML file contains all the necessary data to write the final representation of the sentence in Gloss annotation system.

In the following figure, we present the XML file configuration as the result of our ArabSTS for the sentence " كم الساعة ؟ " (What time is it?). For the translated sentence the interrogative punctuation was deleted and the order of the word was changed.

```xml
<?xml version='1.0' encoding='UTF-8' standalone='yes'?>
<sentence id='13'>
  <word id='0' function='pron_interrog' genre='m' num='s' indication='WH-Q'>
  <word id='1' function='noun' genre='f' num='s' indication='0'>
  <word id='2' function='punc' genre='na' num='na' indication='No-sign'>
</sentence>
<sentenceTranslation id='13'>
  <word id='1' diac='الساعة' function='noun' genre='f' num='s' indication='0'>
  <word id='0' diac='كم' function='pron_interrog' genre='m' num='s' indication='WH-Q'>
</sentenceTranslation>
```

Figure 9. XML file configuration for the sentence " كم الساعة ؟ " (what time is it?)

6. CONCLUSION

This paper presents a description of our proposed system ArabSTS aiming to translate Arabic text to Arabic Sign Language. It covers firstly a presentation of the particularities of the input language of this system, and secondly, a description of the global architecture of our system and the main phases. Furthermore, we have defined the Arabic annotation system trying to limit the problems of standardization of ArSL. In addition to that, this annotation system does not cover only syntax and morphology function of ArSL, but also the movement and location of the sign that we called classifiers. ArabSTS covers the linguistics treatments, the definition and generation of an Arabic annotation system for the Gloss and the classifiers as a transcription system from Arabic text to ArSL.

REFERENCES

[1]   K.Al-Fityani and C.Padden, A Lexical Comparison of Sign Languages in the Arab World, Sign Language Research Conference, Florianopolis, Brazil, December 2006. (2008) R. M. de Quadros (ed.).

[2]   A. Alnafjan, Tawassol Master Project, King Saud University, 2008.

[3]   N. Aouiti, Towards an automatic translation from Arabic text to sign language, in: International Conference on information and communication technology and accesibility (Ed.), Fourth Int. Conf. Inf. Commun. Technol. Access., 2013: pp. 1–4.

[4]   N. Aouiti, M. Jemni, Classifiers in Arab Gloss Annotation System for Arabic Sign Language, the 15th International Conference on Computers Helping People with Special Needs. University of Linz, Austria, 2016.

[5]   N. Aouiti, M. Jemni, For a translating system from Arabic text to sign language, In Proceedings of the Conference Universal Learning Design, Paris 2014. Brno: Masaryk University, 2014. p. 33–38. ISBN 978-80-210-6882-7. (Proceedings of the Conference Universal Learning Design, vol. 4. ISSN 1805-3947.).

[6]   N. Aouiti, M. Jemni, S. Semreen, Arab Gloss Annotation System for Arabic Sign Language, The fifth International Conference of Information and Communication Technology and Accessibility. 2015.

[7]   J. a Bangham, S.J. Cox, R. Elliott, J.R.W. Glauert, I. Marshall, S. Rankov, M. Wells, Virtual Signing : Capture , Animation , Storage and Transmission – an Overview of the ViSiCAST Project Face-to-Face Transactions, Gesture. (2000) 1–7.

[8]   J. Chauché, V. Prince, S. Jaillet, M. Teisseire, Classification automatique de textes à partir de leur analyse syntaxico-sémantique, Proc. TALN. (2003) 11–14.

[9]   S. Cox, M. Lincoln, J. Tryggvason, M. Nakisa, M. Wells, M. Tutt, S. Abbott, TESSA, a system to aid communication with deaf people, Fifth Int. ACM Conf. Assist. Technol. (2002) 205–212.

[10]   O. El Ghoul, M. Jemni, Multimedia courses generator for deaf children, Int. Arab J. Inf. Technol.

6 (2009) 458–463.

[11] C. Hagège, C. Roux, Entre syntaxe et sémantique : Normalisation de la sortie de l ' analyse syntaxique en vue de l ' amélioration de l ' extraction d ' information à partir de textes Mots-clefs – Keywords Résumé - Abstract Analyse syntaxique robuste, (2003) 11–14.

[12] M. and D. Almohimeed, Abdulaziz, Wald, Arabic Text to Arabic Sign Language Translation System for the Deaf and Hearing-Impaired Community, in: EMNLP Second Work. Speech Lang. Process. Assist. Technol. (SLPAT), Edinburgh, UK, Scotl., 2011: pp. 101–109.

[13] A. Othman, M. Jemni, A.T. Hussein, B. Menara, Statistical Sign Language Machine Translation : from English written text to American Sign Language Gloss, Int. J. Comput. Sci. Issues. 8 (2011) 65–73.

[14] The Stanford website, Http://nlp.stanford.edu/software/lexparser.shtml.

[15] A. Pasha, M.Al-Badrashiny and al. MADAMIRA: AFast Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic

[16] M. Mohandes, S. Arabia, Automatic Translation of Arabic Text to Arabic Sign Language, ICGST Int. J. Artif. Intell. Mach. Learn. AIML. 6 (2006) 15–19.

[17] A. Pasha, M. Al-badrashiny, M. Diab, A. El Kholy, R. Eskander, N. Habash, M. Pooleery, O. Rambow, R.M. Roth, MADAMIRA : A Fast , Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic, Proc. 9th Lang. Resour. Eval. Conf. (2014) 1094–1101.

[18] S. Liddell, Grammar, gesture, and meaning in American Sign Language, (2003).

[19] S. Sameer and M. Albanali, قواعد لغة الإشارة, (2009).

[20] L. Zhao, K. Kipper, W. Schuler, C. Vogler, M. Palmer, N.I. Badler, A Machine Translation System from English to American Sign Language - viewcontent.cgi, 1934 (2000) 54–67.