

# SemSignWriting: A Proposed Semantic System for Arabic Text-to-SignWriting Translation

Ameera M. Almasoud, Hend S. Al-Khalifa

Information Technology Department, College of Computer and Information Sciences, King Saud University, Riyadh, KSA.  
Email: ammalmasoud@ksu.edu.sa, hendk@ksu.edu.sa

Received May 26<sup>th</sup>, 2012; revised June 29<sup>th</sup>, 2012; accepted July 10<sup>th</sup>, 2012

## ABSTRACT

Arabic Sign Language (ArSL) is the native language for the Arab deaf community. ArSL allows deaf people to communicate among themselves and with non-deaf people around them to express their needs, thoughts and feelings. Opposite to spoken languages, Sign Language (SL) depends on hands and facial expression to express the thought instead of sounds. In recent years, interest in translating sign language automatically for different languages has increased. However, a small set of these works are specialized in ArSL. Basically, these works translate word by word without taking care of the semantics of the translated sentence or the translation rules of Arabic text to Arabic sign language. In this paper we present a proposed system for semantically translating Arabic text to Arabic SignWriting in the jurisprudence of prayer domain. The system is designed to translate Arabic text by applying Arabic Sign Language (ArSL) grammatical rules as well as semantically looking up the words in domain ontology. The results of qualitatively evaluating the system based on a SignWriting expert judgment proved the correctness of the translation results.

**Keywords:** SignWriting; Arabic Sign Language; Deaf; Semantic Web; Ontology

## 1. Introduction

Sign Language (SL) is an essential communication channel used among deaf people. SL is the native language for deaf people. Also, it's a secondary language for their hearing parents, the hearing children of deaf adults as well as to the hearing deaf educators. Deaf people are facing many difficulties when communicating with other hearing people and in education; this can be attributed to the limited resources of information written in their language. Therefore, an automatic translation system from Arabic text to Arabic Sign Language (ArSL) can help an Arab deaf people to make more information and services accessible to the Arab deaf community in addition to help them learn the spoken Arabic language.

SL basic parts consist of Manual Features (MF) and Non-Manual Features (NMFs). Manual Features are signs performed by one or both hands in different shapes, locations, movements and orientations to represent meaning. While NMFs are features that do not involve hands, however they are used to give meaning, feeling and/or represent the morphological and syntactic markers of a sentence [1].

Previous attempts were carried out to transcribe SL to a written format similar to spoken languages. Stokoe notation [2], HamNoSys [3], Gloss notation [4] and

SignWriting [5] are among such visual representations. By comparing these notations based on a set of features [6], such as: 1) Representation; 2) Language dependency; 3) Usage; 4) Usability for Deaf; 5) Way of Writing; 6) Number of symbols; and 7) NMF, we can find that SignWriting is usually preferred because: 1) it is language independent, which contains large number of basic symbols that can be used to build large number of new symbols; 2) it has a better support for NMF; 3) it is understandable, practical; and 4) it is usable by deaf people in their daily life such as education, communication and reading [6].

Symbols used in SignWriting are pictures that are similar to real world objects, in opposite to Stokoe notation which uses symbols similar to the English alphabet and HamNoSys which uses its own symbols. SignWriting is suitable for all sign languages. International SignWriting Alphabet (ISWA) [7] defines 30 groups of symbols to form 639 base symbols and 35,023 final symbols for better representation and coverage [8]. These symbols describe the form, movement and location of hands, shoulders, fingers and others. It is the first notation system that codifies facial expressions such as eyes, eyebrows, nose, mouth, teeth, tongue, cheeks and breathing.

SignWriting system is now used as a handwriting version of SL and taught to Deaf children and adults in the

world [9].

Given the importance of SignWriting to the deaf community and realizing the fact that few applications were developed to translate text to SignWriting, we propose in this research the development of a novel system that translates Arabic text to SignWriting using semantic web technologies.

To the best of our knowledge, consolidating the domain of SL writing with semantic web is rarely researched. Therefore, the main objective of this research is to investigate the applicability of semantic web technologies namely ontologies to enhance the process of text to SL writing translation. Ontologies provide means of describing entities of an application domain in a well-structured way. Thus, our translation system is limited to the domain of jurisprudence of prayer, because it is a small domain with limited vocabulary and it is really needed for educating Arab deaf Muslims.

In this paper we present SemSignWriting, an experimental semantic translation system for translating Arabic text to SignWriting notation based on ArSL rules and domain ontology. We believe that our proposed system can be the bases for future projects that intend to build Animated ArSL translators.

The rest of the paper is organized as follows. Section 2 presents the problem definition. Section 3 presents previous works in the domain of SignWriting and semantic web technologies for text to sign language translation. Section 4 describes SemSignWriting system design and implementation. Section 5 discussed the results of preliminary experiments carried out to evaluate our system. Finally, Section 6 concludes the paper by discussing our system limitations and future recommended improvements.

## 2. Problem Definition

Previous work in translating Arabic text to ArSL are very few, most of these research worked only on translating word by word and did not take care of the semantics of the translated sentence or the translation rules of Arabic text to Arabic sign language. To resolve this problem, we aimed in our proposed system to enhance previous research in this field by adding an extra layer of semantics while translating Arabic text to ArSL aided by the power of semantic web technologies.

## 3. Related Work

An analysis of previous research in SL translation yielded little work in the area of semantic translation using ontologies and in the field of translation to SignWriting notation. In fact, SignWriting translation is not well-investigated in the literature. Furthermore, previous work in the field of text to Arabic Sign language transla-

tion is also few. This section highlights some previous research that tackles the three domains.

### 3.1. SL and Semantic Web

It was hard to find previous work that uses semantic web technologies to enhance text-to-SL translation. The only reported work was ATLAS [10]. ATLAS is a project for automatically translating from Italian text to Italian Sign Language (LIS). The translation system communicates with the user through a virtual signer: the system takes a text written in Italian language and translates it into a formal intermediate representation of a sign language sentence called ATLAS Written Italian Sign Language (AWLIS). AWLIS sentences are then translated into a character's gestures Animation Language (AL) which describes the way the basic movements are produced and linked. Finally, AL sequences are represented using a 3D representation engine to produce the corresponding animation of the avatar.

ATLAS [11] Linguistic analysis is composed of three steps: 1) deep syntactic analysis of the Italian source sentence; 2) semantic interpretation; and 3) generation of the target LIS sentence.

The Syntactic analysis relies on a morphological dictionary of Italian and on a rule-based grammar to create a dependency tree that represents the syntactic analysis of the source Italian sentence. The Semantic interpretation was built around the weather reports domain ontology. The ontology is used to build a semantic representation of the input sentence, which is then used by the generative process. Basically the system searches for a match between the syntactic trees and the concepts in a domain ontology to find the overall meaning of the sentence which called ontological restriction. Finally, the generation of the target LIS sentences uses the OpenCCG morpho-syntactic generation system. They also convert the ontological restriction into a first-order logic formula to generate LIS sentences.

### 3.2. SL and SignWriting

Prior work tackling the use of SignWriting notation was limited to two key applications themes, which are: "SignWriting editing and writing" and "SignWriting translation".

For SignWriting translation, Ahmed and Seong developed a SL system used for writing and reading text messages in signs as an alternative to Short Message Service (SMS) on mobile phones [12]. The SignWriting system was used to convert text to sign message and sign to text message in two-way communication. The system is usable and beneficial to deaf people and also hearing people to communicate and work with SignWriting.

Similarly, Matsumoto *et al.* [13] developed JSPad sys-

tem to write a Japanese sign language (JSL) using SignWriting. JSPad helps their users to write JSL sentences with SignWriting in shorter time. The system takes a Japanese text then split it into signs, these signs are mapped to SignWriting symbols referring to the JSL dictionary then display them on the screen to permit the users to edit the generated signs then add them to the dictionary.

As for SignWriting editing, SignPuddle is a free web application, developed by Slevinski [14]. It is used to add signs to the SignPuddle dictionary, create SignWriting documents using SignText, send emails in SignWriting using SignMail, and search for signs and for sign language texts by a variety of search-formats, including Search by Words, Search by Signs, Search by Symbols and Symbol-Frequency.

### 3.3. SL and Arabic Language

Still, Arabic research in the domain of developing computerized systems to translate Arabic text to Arabic SL are in their infancy. For instance, Mohandes [15] developed a system to translate Arabic text into Arabic SL. The system has a database to store Arabic dictionary words with the corresponding sign representation video. If the user enters a word that is available in the database then the recorded clip will be shown, if the word is not included then finger spelling is done. Tawassol [16] is another system for translating Arabic text to animated Arabic SL. The system is used as an educational tool. It contains a translator, a dictionary of Arabic words for a set of categories, in addition to a finger spelling editor.

### 3.4. Discussion

Given the different domains SL was used in the previously reviewed work, we can find that there is a research niche for the present study to fill. The idea is to consolidate the different technologies to enhance Arabic SL

translation. In the next section we propose the use of ArSL rules along with ontologies to translate Arabic text to SignWriting.

## 4. System Design and Implementation

As we have seen from the previous section, there was no previous work conducted to implement an Arabic SL translation system that benefit from both semantic web technologies and SignWriting notation. In this section, we describe the design of SemSignWriting system that semantically translates Arabic text to SignWriting.

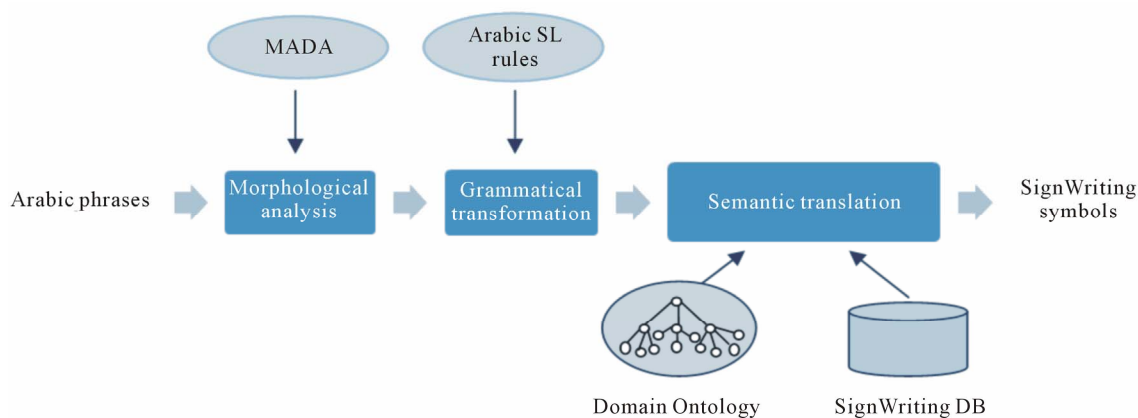
### 4.1. System Desgin

The SemSignWriting system is illustrated in **Figure 1**. The input to the system is an Arabic phrases and the output is a SignWriting symbols (description of these symbols illustrated in the introduction section). System is composed of a set of processes, which are: Morphological analysis, grammatical transformation and Semantic translation.

The Morphological analysis process takes Arabic text as an input and sends each sentence to the Morphological Analysis and Disambiguation for Arabic (MADA) tool for Part of Speech (POS) tagging [17]. (MADA is a system used to address and analyze different natural features for the Arabic language such as: tokenization, part-of-speech (POS) tagging, stemming and lemmatization). Then, the grammatical transformation process takes the previous results as input and applies grammatical Arabic Sign Language rules on each word depending on its POS. Finally Semantic translation takes the result of the previous process and search for each word in the Domain Ontology then show the SignWriting symbol as a result.

### 4.2. System Implementation

SemSignWriting was implemented on the Eclipse IDE using Java programming language to integrate the system



**Figure 1. Proposed system process.**

components. In this subsection we will talk about our system development phases: 1) ontology building; 2) Database (DB) building; 3) user interface design; 4) text processing; 5) grammatical transformation; 6) ontology searching; and 7) sign displaying.

#### 4.2.1. Ontology Building

The ontology was built using protégé editor. Our ontology consists of Arabic WordNet (AWN) ontology as an upper ontology to represent the general Arabic words and our domain ontology to represent jurisprudence of prayer terms, description of these two ontologies are presented in the following two subsections.

##### 4.2.1.1. Upper Ontology

Arabic WordNet (AWN) is a lexical resource based on the Princeton WordNet (PWN) for English language and EuroWordNet (EWN). AWN dictionary developed and linked with the Suggested Upper Merged Ontology (SUMO). AWN consists of 6 tables stored in local database, the tables are: Item, Word, Form, Link, Authorship and Mapping. In our project we did not need all these tables, we just extracted some of them with specific columns in the form of xml file format then imported the generated file to the Protégé editor to build the upper ontology part. The Extracted tables are:

- Item (gloss, itemid, pos): holds information about English and Arabic synsets, synset is a set of one or more synonyms.
- Word (synseid, value, wordid): holds information about words within synsets, for both English and Arabic.
- Form (form\_case, gender, number, person, tense, type, value, wordid): holds information about different forms of Arabic words.
- Link (link1, link2, type): holds links between different synsets or words

The final upper ontology consists of:

- 1) Classes: Item, Word, Form and Link.
- 2) Datatype properties: tense, type, value, person, pos, synsetid, wordidform, case, link1, link2, gender, glositemid and number.

##### 4.2.1.2. Domain Ontology

The domain ontology was built as a basic taxonomy with simple axioms to represents jurisprudence of prayer terms and their SignWriting symbols.

As we can see from **Figure 2**, there are a set of classes in taxonomic (subclass) hierarchy, and a set of properties which link these classes and their instances.

###### 1) Classes:

- Characters “أحرف” class: represent the Arabic alphabets.
- Religion “الدين” class: has a set of 14 sub-classes that

represent jurisprudence of prayer terms.

###### 2) Properties:

- Object properties:
  - is-a.
  - is-kind-of.
  - is-synonym-of.
- Datatype properties:
  - Signid.
  - Synonym.
  - Value.

###### 3) Individuals:

- Characters “أحرف”: 29 instances “the Arabic alphabet”.
- Worship “العبادة”: 54 instances, which are: “دين/Religion-سجود/Prostration-ركوع/Bowing-تكبير/Takbeer-أركان/Pillars-شهادتين/The Testimony Of Faith-صلاة/Prayer-زكاة/Zakat-صوم/Fasting-حج/Hajj-ركن/Pillar-واجب/Duty-إفعية/Act-تأمين/Ameen-إستفتاح/Opening-بسملة/Basmalah-مفروضة/Impose-عبادة/Worship-قولية/Sunnah-سنة/Annuls-مبطلات-Five/خمس-عين/Obligation-فرض-فجر/Fajr-رباعية/Four-ظهر/Zuhr-عصر/Asr-نية/Intention-قيام/Standing-مغرب/Maghrib-عشاء/Isha-كفاية/kifaayah-الأضحى/Nafilah-نافلة/Eid al-Fitr-عيد الفطر-Eid-أحدي/Pray forrain-استسقاء/Weter-وتر/Tahajud-تهجد-Adha-تحية/Greeting-كسوف/Eclipse-جنزة/A Fune al-Taraaweeh-تراويح-Rawateb-إستخارة/Duha-ضحى-مسافر/Patient-مريض-أهلا لأعداء/Exempted People-خوف/Fear-الإسلام/Islam-دعاء/Du'aa”.

##### 4.2.1.3. Linking Domain Ontology to Upper Ontology

Our domain ontology was linked under the Form class in the AWN ontology, thus the Characters “أحرف” and Religion “الدين” classes and their sub-classes became sub classes of the Form class, this is done by importing the domain ontology under the form class, see **Figure 3**.

#### 4.2.2. Database Building

We used Microsoft SQL Server Management Studio Express to build a small DB consisting of single table called signs, which holds the sign IDs and path of each sign symbol.

#### 4.2.3. User Interface

We developed a simple Arabic user interface **Figure 4** that allows users to input an Arabic text in the input area then press the translation button to show the result of the translation process as a set of signs in the output panel.

#### 4.2.4. Text Processing Process

The text processing consists of a sequence of steps which include:

- 1) Filtering the input text from any character other than Arabic letters, spaces and dots.
- 2) Replacing any of these letters “؟”/“E”, “؟”/“A” or

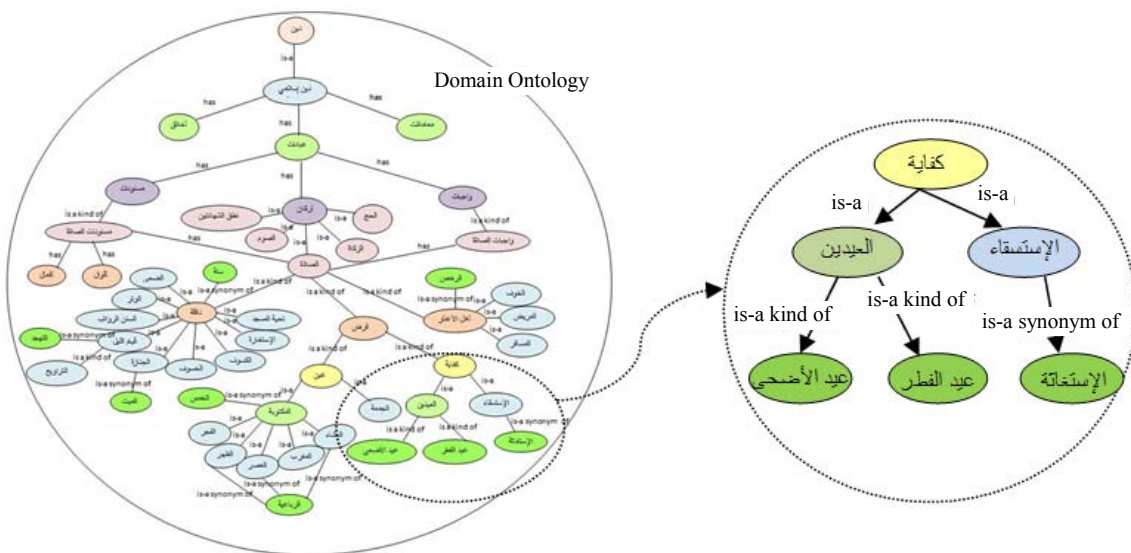


Figure 2. Example of a small part of the domain ontology.

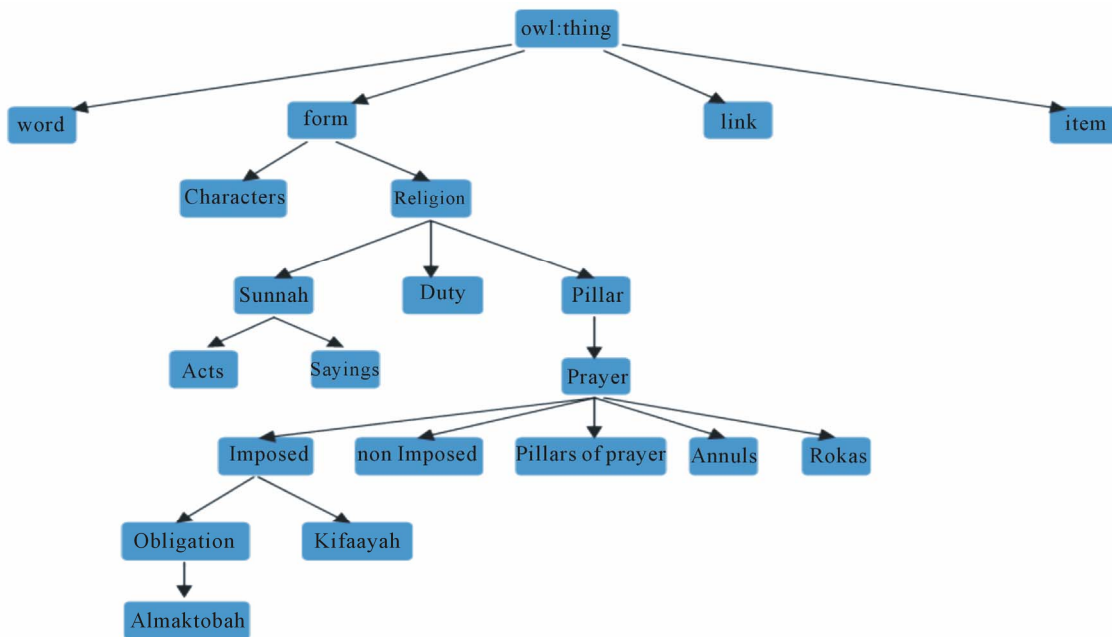


Figure 3. Final ontology.

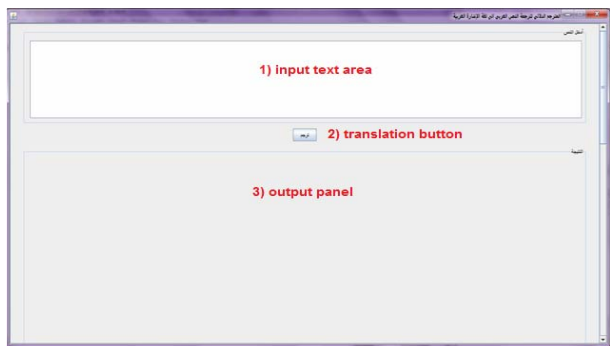


Figure 4. SemSignWriting interface.

“أ” with the letter “ا” before any processing, because words containing any of these letters may be written in many different forms.

3) Breaking the input text into a set of sentences, then writing these sentence in a text file “one sentence per line” to be used by the Morphological Analysis and Disambiguation for Arabic (MADA) program.

4) Invoking MADA to analyze the text file, MADA program is operated through Cygwin “Unix like environment”.

5) Reading the result of MADA program and sending each word with its MADA result features: “part of

speech (POS), gender and number” to the search method if and only if a word is not a preposition, relative pronoun, abbreviation, punctuation, punctuation or interjection.

#### 4.2.5. Grammatical Transformation Process

Experts point out that the linguistic structure of the indicative sentence are in this form (Subject, Verb, Object), but deaf people are using another linguistic structures in the form of (Verb, Subject, Object) or (Object, Verb, Subject). However, the most commonly used linguistic structure is (Subject, Verb, Object) [18]. Therefore, in the grammatical transformation process, the Arabic text that is tagged with POS (from the previous process) is cleaned from any preposition, relative pronoun and then its linguistic structure is reordered accordingly.

#### 4.2.6. Ontology Searching process

The search process consists of the following steps as illustrated in “Figure 5”:

- 1) Searching the ontology for the sign ID of the word with specific POS, gender and number.
- 2) If the word was not found, then search for it in the word’s synonyms.
- 3) If the word’s synonyms was not found also, then it will be given a zero id, words with zero id will be finger spelled.
- 4) Sending each word with its ID to the display method.

#### 4.2.7. Signs Display Process

In order to display the signs of the translated words, the following steps are carried out in the display process:

- 1) A connection to the signs Database is established and a query with the word is setup.
- 2) If a word has a sign, then retrieve its sign path from the DB.
- 3) If a word does not have a sign, then retrieve its character signs paths from the DB for finger spelling.
- 4) Finally invoke the display image method to show each sign symbol.

## 5. Experimental Results

In order to evaluate our system, two qualitative evaluation strategies were followed: the first one is done by translating a set of sentences automatically then asking a domain expert to check their accuracy. The second experiment was done based on comparing the translation result of our system with the expert’s manual translations for a number of sentences. We have only one expert in this field thus the evaluation was conducted on a small data set.

### 5.1. Experiment #1

This experiment was performed by translating a sample composed of 46 sentences automatically using our system then asking a domain expert to check their correctness. The results of the evaluation shows that the expert marked all sentences as correct translation even though we have some words that were finger spelled.

As we can see from **Table 1**, each word in the sentence “النِّيَّةُ الْقِيَامُ الرَّكْعَةُ وَالسُّجُودُ مَنَارُ كِتَابِ الصَّلَاةِ”/“Intention, Standing, bowing and prostration are the pillars of prayer” has a corresponding SignWriting symbol however, the word “خمسة”/“Five” in the second sentence does not have a

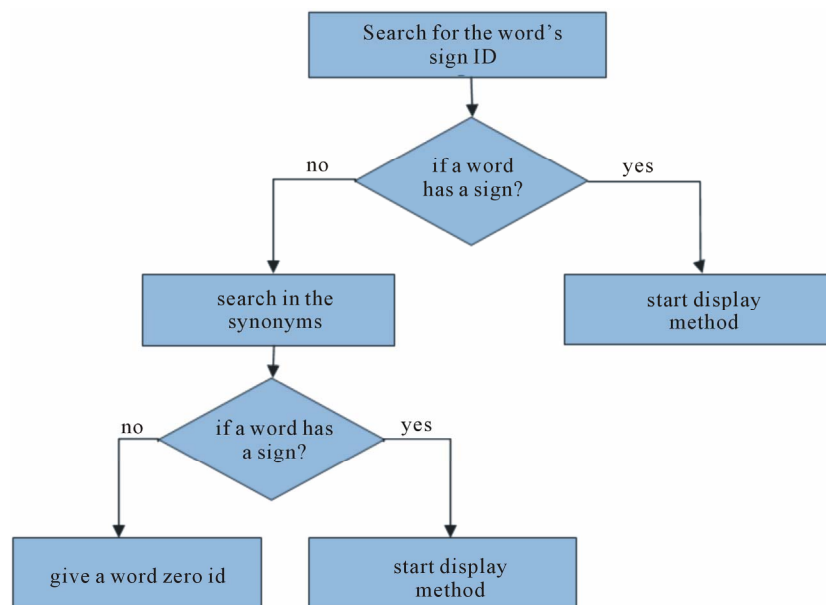


Figure 5. Ontology searching steps.

SignWriting symbol in the Domain ontology therefore it was finger spelled. This result can be attributed to a limitation in our system as we will discuss in Section 5.

Our system was also evaluated to check the consistency of the synonyms, as shown in Table 2. Both sentences "صلاة الإستسقاء سنة"/"Pray for rain is Sunnah" and "صلاة الإستغائة سنة"/"Pray for rain is Sunnah" were resolved to the same SignWriting symbol because the word "الإستسقاء" and "الإستغائة" were marked synonyms in our domain ontology, so that they used the same SignWriting symbol.

5.2. Experiment #2

The second experiment was conducted by comparing the translation results of four Arabic sentences translated by our system against the expert’s manual translation. The result obtained from this experiment showed that the

result of our system translation is correct and equal to the expert’s manual translation result, as shown in Table 3. However, in the second sentence our system finger spelled one of the words because it was out of scope in our domain ontology, as shown in Table 4.

6. Conclusions, Limitations and Future Work

In this paper, we presented a proposed semantic translation system for Arabic text to SignWriting using a morphological analyzer, applying grammatical ArSL rules and performing semantic lookup that replaced each word by its SignWriting symbol using domain ontology. Our proposed system was limited to the jurisprudence of prayer domain and used the AWN ontology as an upper ontology. The novelty of our system relies on supporting the process of Arabic text to SignWriting translation with a layer of semantic technologies using ontologies.

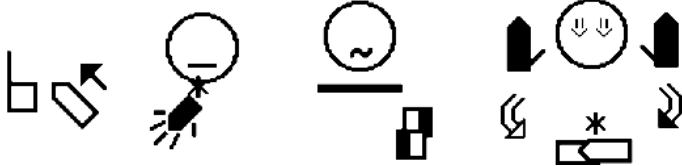
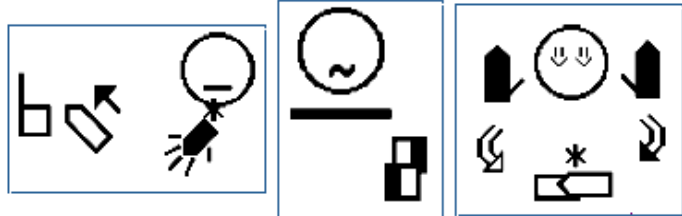
Table 1. Sample of the evaluation result for two different sentences.

النية والقيام والركوع والسجود من أركان الصلاة "Intention, Standing, bowing and prostration are the pillars of prayer"	
(a) A sentence with all signs found in the domain ontology.	
أركان الصلاة خمسة "the pillars of prayer are Five"	
(b) A sentence with the word "خمسة" being finger spelled because of the lack of a corresponding sign writing symbol in the domain ontology.	

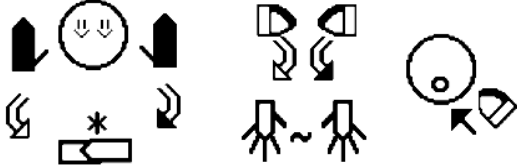
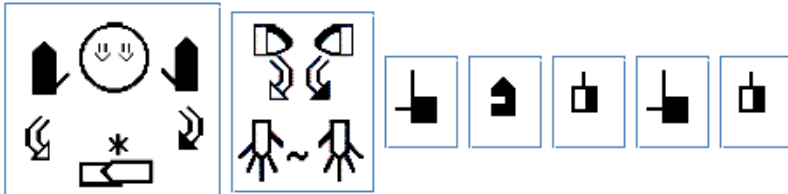
Table 2. Sample of the evaluation result for two synonym sentences.

صلاة الإستسقاء سنة "Pray for rain is Sunnah"	
صلاة الإستغائة سنة "Pray for rain is Sunnah"	

**Table 3. Translation result match: SemSignWriting vs Expert.**

	صلاة الجنّارة سنة / "Funeral prayers is Sunnah"	Result
Expert translation		
System translation		Exact match

**Table 4. Translation result mismatch: SemSignWriting vs Expert.**

	الأكل من مبطلات الصلاة / "Eating is invalidate the prayer"	Result
Expert translation		
System translation		Translation is correct but the word "الأكل"/"eating" was finger spelled in our system translation because it was out of the domain ontology scope.

Given the correctness of the translation results, yet, our system suffers from the following limitations:

- 1) The AWN upper ontology does not cover Arabic numerals.
- 2) The domain ontology does not cover all concepts in the domain; it has only 54 instances.
- 3) We did not obtain all words' signs, which resulted in finger spelling some words.
- 4) We have only one expert in this field thus the evaluation was conducted on a small data set.

Based on the previous limitations inherited in our system, we suggest some potential extensions that can enhance our system performance and results, such as:

- 1) Decode SignWriting symbols to sequences of scripted animation commands.
- 2) Improve the system response time by finding a faster morphological and analysis tools for Arabic language instead of MADA.

Expand the domain ontology to cover more words and concepts.

## 7. Acknowledgements

The authors are very thankful to Dr. Mohammed Abu-shaira the expert in SignWriting translation for evaluating our system results.

## REFERENCES

- [1] P. V. Paul, "Language and Deafness," Jones & Bartlett Learning, Burlington, 2008.
- [2] J. Martin, "About the Stokoe System," 2010. <http://www.signwriting.org/forums/linguistics/ling006.html>
- [3] "About the HamNoSys System," 2011. <http://www.signwriting.org/forums/linguistics/ling007.html>
- [4] A. Almohimeed, M. Wald and R. Damper, "An Arabic Sign Language Corpus for Instructional Language in School," 2010. <http://eprints.ecs.soton.ac.uk/21106/>
- [5] M. Abu Shaira, "The Effect of SignWriting on the Achievement and Vocabulary Acquiring for Deaf Student



- at Al-Amal School for Deaf in Amman City,” Ph.D. Dissertation, Jordan University, Amman, 2007.
- [6] A. M. Almasoud and H. S. Al-Khalifa, “A Proposed Semantic Machine Translation System for translating Arabic text to Arabic sign language,” *The 2nd Kuwait e-Services and e-Systems Conference*, New York, 5-7 April 2011. doi:10.1145/2107556.2107579
- [7] V. Sutton, “SignWriting,” 2010. <http://www.signwriting.org/>
- [8] R. F. de Brito and A. T. C. Pereira, “A Model to Support Sign Language Content Development for Digital Television,” *2009 IEEE International Workshop on Multimedia Signal Processing*, Rio De Janeiro, 5-7 October 2009, pp. 1-6. doi:10.1109/MMSP.2009.5293266
- [9] S. Morrissey, “Data-Driven Machine Translation for Sign Languages,” Ph.D. Thesis, Dublin City University, Dublin, 2008.
- [10] L. Lesmo, A. Mazzei and D. Radicioni, “Linguistic Processing in the Atlas Project,” SLTAT 2011, Berlin, 2011.
- [11] L. Lesmo, A. Mazzei and D. P. Radicioni, “Linguistic Descriptions in Ontology-Based Machine Translation,” In: B. Kokinov, A. Karmiloff-Smith and N. J. Nersessian, Eds., *European Perspectives on Cognitive Science*, 2011.
- [12] A. S. Ahmed and D. S. K. Seong, “SignWriting on Mobile Phones for the Deaf,” *Proceedings of the 3rd International Conference on Mobile Technology, Applications & Systems—Mobility*, Bangkok, 25-27 October 2006, p. 28.
- [13] T. Matsumoto, M. Kato and T. Ikeda, “JSPad: A Sign Language Written Tool Using SignWriting,” *Proceedings of the 3rd International Universal Communication Symposium*, Tokyo, 3-4 December 2009, p. 363. doi:10.1145/1667780.1667855
- [14] V. Sutton and S. Slevinski, “SignPuddle Reference Manual,” 2010. <http://www.signbank.org/signpuddle/>
- [15] M. Mohandes, “Automatic Translation of Arabic Text to Arabic Sign Language,” *AIML 06 International Conference*, Sharm El Sheikh, 13-15 June 2006, pp. 71-76.
- [16] A. Alnafjan, “Tawassol,” Master Project, King Saud University, Riyadh, 2008.
- [17] R. Roth, O. Rambow, N. Habash, M. Diab and C. Rudin, “Arabic Morphological Tagging, Diacritization, and Lemmatization Using Lexeme Models and Feature Ranking,” *Proceedings of ACL-08: HLT, Short Papers*, Columbus, 15-20 June 2008, pp. 11-120 <http://www.aclweb.org/anthology/P/P08/P08-2030>
- [18] S. Samreen and M. Benali, “The Rules of United Qatari Arabic Sign Language,” 2009.