A Humanoid Robot as a Translator from Text to Sign Language

Siham Al-Rikabi, **Verena Hafner**

Cognitive Robotics Group, Department of Computer Science
Humboldt-Universität zu Berlin
Unter den Linden 6, 10099 Berlin, Germany
{alrikabi, hafner}@informatik.hu-berlin.de

Abstract

This paper presents a new algorithm to translate Arabic texts into American Sign Language (ASL). This task is performed by Nao, a humanoid robot. In an intensive study conducted on the complete ASL dictionary, the challenges of using a humanoid robot for the translation become evident. The results of this study demonstrate which part of the sign language could be performed by the robot and how to figure out the common denominators between Arabic and ASL grammars in order to create good mapping for an intermediate parse tree.

Keywords: Nao, humanoid robot, ASL, Arabic, translation algorithm

1. Introduction

Deaf people everywhere still share a long-standing problem. Communication with hearing people who do not understand sign language is not possible. Therefore auto-translation from text into sign language could be used to make their lives easier. In this field, several serious attempts have been made to create interaction between deaf and hearing people by using advanced technology. However, there were not many practical solutions until now.

An attempt that could be considered as the first powerful solution to this problem is an invention that Ryan Patterson made in 2000 called the American Sign Language Translator (Parton, 2005). The idea behind his system is that deaf people should wear an electronic glove that sends a radio frequency signal to a PC connected to a portable display device. When the deaf person hand spells a word, the letters appear on the display device sequentially. Before Patterson's invention, in 1977 SWRI in the USA (Jaffe, 1994) designed and developed a robotic finger-spelling hand to help deaf and blind people to obtain new information through a simple procedure whereby the deaf and blind user puts his/her hand on the robotic hand and somebody types a message on a keyboard that is connected to the robotic hand through an electrical logic circuit. In 1985, Dexter, which is a large robotic finger-spelling hand that stands on a box, was invented and developed at Stanford University.

Some studies focused on the learning aspect because many deaf people do not know how to read a written text in the native language of their country. Learning of any scientific field, for instance, will be a hard challenge for deaf people. One of the most crucial projects is called TEAM (Translation from English to ASL by a Machine) by Zhao and others in 2000 (Zhao et al., 2000). They designed a human model avatar in order to perform the signs of ASL which have been matched with the English entry text. They also generated an intermediate representation from the entry text by doing several steps that are: analyze the entry word order and find out the most appropriate sign, then generate ASL glosses to find a good mapping from English to ASL by using Synchronous Tree-Adjoining Grammar (STAG), and finally produce ASL signs to be performed by the human avatar.

In 2008, Halawani introduced a translation system based on a mobile device from Arabic text to Arabic sign language for the deaf users who could not read Arabic text. Using this system a deaf person can write and send the Arabic text to the suggested system through a mobile device by using a Web browser and some protocols to transfer data, and the system translates the entry text into animated signs which appear on the monitor of the mobile device (Halawani, 2008).

However, even though all those projects are very useful in the communication with hearing people aspect, they could not be in use as much as expected for many reasons tied to those inventions such as cost, size, time of processing, and stability. Regarding the research of translation from natural language to sign language, it is still at the very beginning. Also the employment of robots to perform signs of sign language was limited to one type of a robot – the hand robot – that is utilized for the sign language finger spelling purpose.

In this paper, the most important aspects of this research field will be illustrated, and a new translation algorithm from Arabic text to ASL, based on a study of the complete ASL dictionary, will be introduced. Finally, we will explain how the ASL signs can be performed by a humanoid robot and discuss the prerequisites, challenges and limitations.

2. Text and Sign Languages Used

In this section, the specific languages utilized in our translation system are presented to explain the common denominators between them. The choice for text entry is Arabic because of previous studies and experience of the first author, as well as some grammatical features of this language presented in the next subsection. The choice for the sign language is American Sign Language (ASL) because of its simplicity and popularity.

2.1. Arabic Language

Arabic is a major member of the Semitic language family. After the appearance of Islam, it had become the official language of the old empire of Islam because it is the language of Qur'an, which is the holy book of Islam.

Nowadays over 300 million people are Arabic native speakers, and more than 50 million people speak Arabic as their second language (Boullata and DeYoung, 1997). Arabic speech consists of verb, noun (nouns, pronouns, adjectives and adverbs), and particles (preposition, articles, conjunctions, and interjections). Arabic sentences classify into two main types; nominal sentence (the sentence starts with a noun), and verbal sentence (the sentence starts with a verb). The structure of a nominal sentence is “Topic-Comment”; Topic could be a word or a phrase and Comment is the predicate of a nominal sentence (Dehdah, 1996). The Arabic language holds several challenges, one of them being that some letters of the words may be converted to another letter during the morphological analysis of the word. Another challenge is that modern Arabic text has been written without diacritics (short vowels). On the other hand it boasts mighty features which assist in natural language processing, such as deriving of all the words which project the same theme from one root. For instance, the theme of writing “كتابة” includes the verb in all tenses, the subject of writing, the object of writing, and the adjective. All nouns which could be derived from the identical root “ك ت ب” depend on the morphological pattern (MP), whereas English requires many roots to express the same theme as shown in Table 1. The Arabic MP presents syntax, semantics, and morphological features (Alrikabi, 2000).

Word	Meaning	MP	Comment
يكتب	He writes	يفعل	Masculine (M)-singular subject
تكتب	She writes	تفعل	Feminine (F)- singular subject
نكتب	We write	نفعل	M/F–plural speakers subject
أكتبُ	I write	أفعلُ	Singular speaker subject
يكتبان		يفعلان	M-dual subject
يكتبون	They write	يفعلون	M/F plural subject
تكتبين	You write	تفعلين	F-singular addressees subject
كتب	He wrote	فعل	M-singular subject
كتبتُ	I wrote	فعلتُ	Singular speaker subject
كتبتْ	She wrote	فعلتْ	F-singular subject
كتبتا		فعلتا	F-dual subject
كتبوا	They wrote	فعلوا	M-plural subject
أكتبْ	write	أفعلْ	M-singular addressee subject
أكتبي	write	افعلي	F-singular addressee subject
أكتبا	write	أفعلا	M/F-dual addressee subject
اكتبن	write	افعلن	F-plural addressees subject
كتابة	Writing	فعالة	
كاتب	Writer/Author	فاعل	
كتاب	book	فعال	
مكتبة	library	مفعلة	
مكتب	office	مفعل	

Table 1: This table shows some Arabic words that can be derived from one root “ك ت ب” and also their English translation.

2.2. American Sign Language (ASL)

ASL is the official language of the American deaf community. It holds its own phonology, morphology, syntax, and pragmatics, which are completely different than the spoken English. ASL sentences embrace two structures; the first sentence has a “Topic-Comment” structure, while others adopt a “Time-Topic-Comment” or “Time-Comment” structure. The topic presents the subject of the sentence, while the comment is the predicate of the sentence. The verb of an ASL sentence employs a simple form (present), and it is appended with special words that indicate the time of the event, such as “just”, “now”, “tomorrow”, etc. Also nouns are used in their singular form, in order to represent the plural, the countable noun is appended with a number. The ASL sentence could be a nominal sentence (no verb is utilized in the sentence like in Arabic grammar) because there is no verb for to be in ASL as in English (Humphries and Padden, 1992).

3. Nao - the Humanoid Robot Utilized in this System

Nao is a humanoid robot with various sensors and actuators. The standard Nao robot used in this study has 21 degrees of freedom (DOF). DOF refers to the total number of independent rotations of the joints. This robot is limited in its ability to perform sign language since it does not have any DOF in its hands and no human-like emotional expressions. We will investigate the possibilities and limitations of the robot in the next section and compare it with two other robot platforms.
The academic version of Nao (*AcNao)* has 25 DOF. *AcNao* differs from the standard Nao only by 4 DOF in the robot hands. Figure 1 presents a Nao and its features (Aldebaran, 2011). An ideal robot, which can perform all the signs of the ASL dictionary including facial expression, should have 80 joints and 135 degrees of freedom. These numbers include joints and muscles in the robot face to present facial expression according to a human model avatar, which is presented by Zhao and others (Zhao et al., 2000). There is currently no *ideal robot* on the market, therefore we tend to process as many signs of ASL as possible according to the current limitations. Although certain advanced robots (*AdRobot)*, such as *Robonaut R2*[1] and Mobile Humanoid *Rollin' Justin* [2] have the ability to express all the signs except for facial expressions, they are not easily available for researchers.

4. Analysis of the ASL dictionary

Due to the lack of information in the computational linguistics of ASL, we did an empirical study based on the ASL dictionary, in which:

- The complete ASL dictionary, which includes 6719 words, has been rewritten as a text file (available for researchers on the website of the authors).
- Each single word video of the dictionary has been viewed and classified depending on whether the

[1] *Robonaut R2 from Nasa, for further information see the link:* http://robonaut.jsc.nasa.gov/default.asp
[2] *Rollin' Justin from DLR, see the link below for further information;* http://www.dlr.de/rm/en/desktopdefault.aspx/tabid-5471/

standard Nao, AcNao or AdRobots could perform those signs, the needs for facial expressions, and the complexity of the word signs.

The following Table 2 shows the results of this study (details about facial expressions are available in the full dictionary).

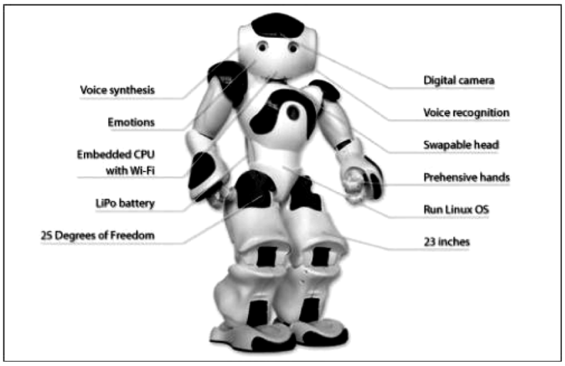

Figure 1: This figure describes the Nao robot in terms of its joints, sensors, etc.

Percentage of ASL signs that could be performed	*Nao*	*AcNao*	*AdRobot*
Fully	4%	40%	82%
Partially	18%	46%	18%
In total	22%	86%	100%

Table 2: The percentage of signs that could be performed by the different humanoid robot platforms.

The study demonstrates that 1227 signs from the ASL dictionary require facial expressions, which is only 18% of the ASL dictionary. This means that even without an ideal robot, 82% of the dictionary can be performed with *AdRobot*.

Another factor that is crucial to represent the signs is their complexity. Complexity has been classified into three levels *C1, C2* and *C3.* It depends on the number of nodes *N* in the *MotionNet*, a topological representation of the motions to be performed by the robot, see Table 3.

Percentage of ASL signs that could be performed	*C1:* $1 < N \leq 6$	*C2:* $6 < N \leq 12$	*C3:* $N > 12$
In total	38%	50%	12%
Nao	11%	10%	1%
AcNao	34%	43%	9%
AdRobot	16%	26%	9%

Table 3: Here, the number of possible signs separated by complexity is shown based on our study.

This study clearly highlighted some facts, which were completely or partially ignored by computational linguistics and deaf studies. These could help formalize ASL in order to process it automatically, and not least, support researchers who are interested in this field.

- The ASL dictionary holds a number of words 8 times less than the Oxford English dictionary.
- Most of the signs that need facial expression are tied to *emotions* and *weather* themes.
- The words, which have the same meaning, have the same sign for example *"author"* and *"writer"*.
- Finger-spelling is utilized for names, and when a new product is released.
- All interrogative tools need facial expression. For instance *"why"*, *"where"*, etc.
- Pronouns that are related to the same person or people, have the same sign (for instance I, me, mine, my) and we can figure out which one should be utilized depending on the structure of the sentence. The pronouns are presented by pointing to the meant person or people, which would be problematic with an avatar.
- A verb in ASL could be considered as the root of all words that talk about the same theme. For example the word *"painter"* consists of two signs, the first one is the sign of the verb *"paint"* and the sign of the noun *"person"* comes secondly.
- In most cases, the noun of ASL has the same sequence of verb signs, but the sign starts from a different position. An example is *"instruct"* and *"instruction"*.
- The negation word comes at the end of the sentence, which is similar to German grammar.
- Some signs are represented only by facial expression such as the word *"shiver"*.

5. Translation Algorithm

We first built a translation algorithm for simple patterns of Arabic sentences whereby each sentence matches one *MotionNet*. Then, the algorithm was improved based on our study of the ASL dictionary to generalize and to include the most frequent patterns of Arabic sentences, and to be implemented later on an *AdRobot*. Before the algorithm is presented, the previous work has to be clear. In 2000, we came up with a new concept whereby our semantics system was based on a lexicon consisting of only roots instead of words[3]. Besides the lexicon of roots, a dependable morphological analyzer was built for most morphological patterns of Arabic in order to save time and memory that were crucial factors. Each word has obtained fully certain, and uncertain features[4], which are inferred during the run-time of the morphological analyzer. The uncertain features are completed automatically during the processing of the syntax analyzer. Those features are crucial to build the parse of the entry sentence as well as to understand the full meaning of the sentence. The syntax analyzer adopts the *Dependency Grammar* concept in order to obtain the required result. An example in figure 2 demonstrates the output of the entry sentence "كتب محمد الدرس" for the morphological and syntax analyzers. The output of the morphological analyzer is non-connected nodes of a tree. Each word has one node except in the case of ambiguity whereby the word holds two nodes. The entry text has no diacritics when some morphological patterns of verbs look like those of a noun without diacritics, thus it triggers this ambiguity. To solve this problem, the

[3]Lexicon of words needs additional information about semantics, syntax, morphological, and phonological features have to be connected with it

[4]These features are semantics, syntax, morphological, and phonological features

morphological analyzer provides two nodes as *verb* and *noun*, then the syntax analyzer decided which one is the correct one depending on the grammar pattern and the inferred features of the word. As shown in figure 2, the features of the verb are different from those of the noun (Alrikabi, 2000).

Input sentence: كتب محمد الدرس= "katab Mohammed aldars"

The output of the Morphological Analyzer are:

Word	MP	Subject	Object	Declension	Tense	Root
كتب	فعل	Not yet	Not yet	Non-inflect	past	كتب

Word	MP	Pre-word	Plurality	Gender	Definiteness	Declension	Genitive	Root
كتب	فعل	none	broken	male	no	nominative	possible	كتب
Word	MP	Pre-word	Plurality	Gender	Definiteness	Declension	Genitive	Root
محمد	مفعل	كتب	singular	male	no	nominative	possible	حمد
Word	MP	Pre-word	Plurality	Gender	Definiteness	Declension	Genitive	Root
الدرس	فعل	محمد	singular	male	yes	nominative	No	درس

The output of the Syntax Analyzer are:

Word	MP	Subject	Object	Declension	Tense	Root
كتب	فعل	محمد	الدرس	Non-inflect	past	كتب

Word	MP	Pre-word	Plurality	Gender	Definiteness	Declension	Genitive	Root
محمد	مفعل	كتب	singular	male	no	nominative	possible	حمد
Word	MP	Pre-word	Plurality	Gender	Definiteness	Declension	Genitive	Root
الدرس	فعل	محمد	singular	male	yes	accusative	No	درس

Figure 2: This figure shows the result of the morphological and syntax analyzers.

These analyzers and the lexicon of roots of Arabic were employed by the suggested translation algorithm. At the same time, the lexicon of roots of Arabic inspired us to represent the lexicon of ASL as a lexicon of roots (a root in ASL has the form of a verb in the present tense, or a noun which does not derive from a verb). That process assists to find a clear mapping from Arabic to ASL in addition to reduce the size of the ASL dictionary that is built for the purpose of translation. For instance, the verb "*paint*" can be represented in Figure 3 by the pictures 1-5 as well as the pictures 6-7 demonstrate the sign of the word "*person*", whereas the whole figure demonstrates the word "*painter*". For the current example, we kept in our ASL dictionary the words: "*paint*" and "*person*", but the word "*painter*" could be concluded from the features which are related to the word after the morphological and syntax phases have been done. Thus, the structure of our dictionary is root→ signs. All signs could be represented by a *MotionNet* which consists of *KeyFrames* and *Transitions*. A *KeyFrame* represents the position of the joints of the robot while a *Transition* represents the time between two *KeyFrames.* Each *KeyFrame* expresses the values of the joints of a robot for *Head, Shoulder, Elbow, Hip, Knee* and *Ankle*, in order to obtain the required positions of joints to represent the sign of a word. Figure 4 shows the same word *painter* that is performed by *Nao.* A *MotionNet* could be created through *the Motion Editor*[5]. The technique of our translation algorithm is presented by the following sequence of steps. Firstly, it checks the entered text to prove it holds no unknown symbols. Secondly, the morphological analyzer tests each word of the text to find the root and infer the features depending on the morphological pattern. Here, we should mention that each morphological pattern holds an individual segmentation algorithm according to the length of the word as well as the expected letters which are probably appended to the original root to compose the word. Thirdly, the syntax analyzer fills in the empty fields of the features and corrects the uncertain features, deletes the incorrect nodes, and at the end of its processing, builds the parse tree of the sentences according to the grammar pattern. Fourthly, the tree of the Arabic sentences is converted to the compatible tree of ASL. To achieve that, we observed the similarity between Arabic and ASL grammar, especially in case of *the nominal sentence,* the tree of the ASL sentence holds the same structure as Arabic. Otherwise, the structure of an Arabic sentence would be converted to the form "Time-Comment" except for the first sentence of each paragraph, which has the form "Topic-Comment". Finally, the robot performs the signs of ASL by tracking the sequence of the ASL tree, and checking each word of the ASL tree in the suggested dictionary of ASL. When the word holds the same form of the root, the robot performs it directly. Otherwise the algorithm will utilize the additional acquired features, which are connected to the word during the morphological and syntax analyzers, to compose the full sign of the word, and later perform it by the robot. Figure 5 demonstrates the suggested translation algorithm.

[5] Motion Editor has been built by the Nao Team Humboldt (NaoTH)

6. Conclusions and Future Work

Using a humanoid robot to perform sign language instead of an avatar has several advantages; a humanoid robot can physically and naturally interact like a human being, and it is mobile. This is relevant for an intuitive interaction that includes pointing gestures and joint attention. The employment of a humanoid robot as a medium to solve the communication problem of deaf people could be a good motivation for the robotics industry as well as for researchers in our field. Our translation algorithm is generalizable and can be comfortably applied to other languages that have the same grammar. It can also be employed on a different humanoid robot. We showed that even with a limited robotic platform, a certain percentage of the signs in the ASL dictionary could be performed non-ambiguously.

We are currently working on video-based sign language recognition in order to realize human robot interaction through a general ASL conversation based on our study of the ASL dictionary. In addition, we plan to solve the problem of facial expressions by finding a new representation that can be made with the Nao.

Figure 3: The images are taken from a video sequence of a person performing ASL to represent the word "painter" (**© 2004, www.Lifeprint.com. Used by permission.**)

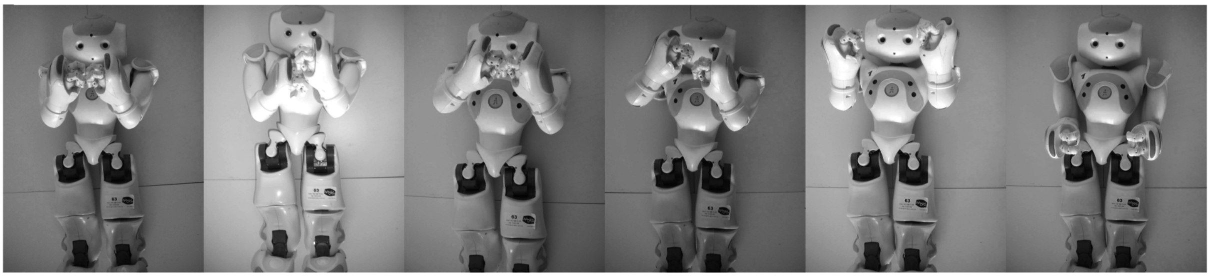

Figure 4: The images present a video sequence of Nao performing ASL to represent the same word "painter".

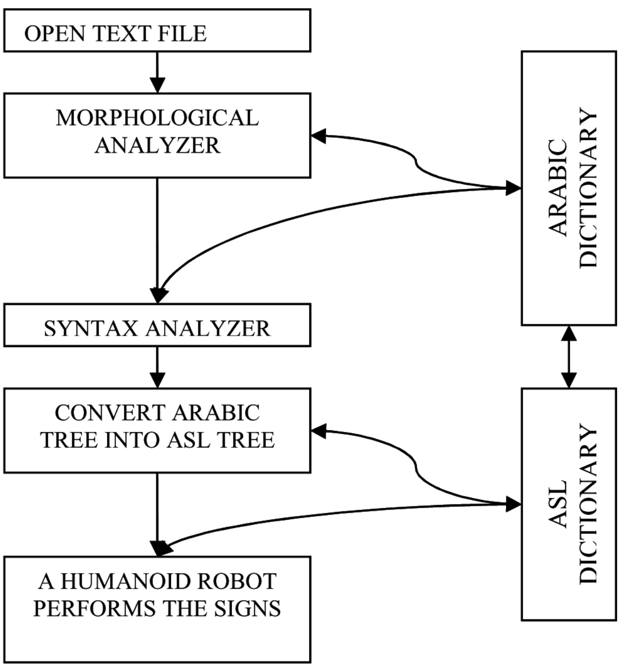

Figure 5: Diagram of the translation algorithm.

Acknowledgments

We thank the Nao Team Humboldt for their helpful software tools and their support as well as Sasa, Lorenzo and Maithim for their valuable comments on the manuscript.

References

Alrikabi, S. (2000). *A Real-Time Intelligent System to Understand Arabic Sentences*. Master thesis. Basra University, Science College, Computer Science department.

Boullata, I. and DeYoung, T. (1997). *Tradition and modernity in Arabic literature*. Fayetteville: The University of Arkansas Press.

El-Dahdah, A. (1996). *A Dictionary of Arabic Grammar in Charts and Tables.* Lebanon Library Publisher.

Halawani, S. (2008). *Arabic Sign Language Translation System On Mobile Devices*. IJCSNS International Journal of Computer Science and Networks Security, VOL.8 No.1, January 2008.

Humphries, T. and Padden, C. (1992). *Learning American sign language*. Englewood Cliffs, N.J: Prentice Hall.

Jaffe, D. (1994). *Evolution of mechanical fingerspelling hands for people who are deaf-blind.* Journal of Rehabilitation Research and Development. Vol. 31. No. 3, pp. 236-244.

Parton, B. (2005). Sign Language Recognition and Translation: *A Multidisciplined Approach from the Field of Artificial Intelligence.* Journal of Deaf Studies and Deaf Education Advance Access, Published by Oxford University Press.

Zhao, L., Kipper, K., Schuler, W., Vogler, C. and Palmer, M. (2000). *A Machine Translation System from English to American Sign Language*. Lecture Notes in Computer Science, Volume 1934, Envisioning Machine Translation in the Information Future 4th Conference of the Association for Machine Translation in the Americas, pp. 54-67.

Aldebaran. (2011). *Technical Specifications* from: http://www.aldebaran-robotics.com/en/node/1166. Access date: May 1, 2011.