

Machine Translation of Arabic Verb Sentences into English

Zainab Abraham Alfahad Jenan Abd Al_Kareem Abd Al-Hasan
Computer Science Department Internet Center
College of Science - Basrah University

Abstract

Arabic is one of the most complex languages as it has characteristics that set it apart off other languages . These characteristics have been a rich resource of study for researchers in the field of Arabic , computational linguistics . This research is considered one of the source as it is expected to play a significant role in the study of Arabic , computational linguistics .

Text translation especially instant translation is regarded one of the demands of the present time imposed by the technology of information exchange through the internet and its necessity in the media and conferences.

The research focuses on the design of model for a machine translation from Arabic to English which translates some Arabic verbal sentences of different orders . The proposed model consists of three main parts which one analysis , dictionary , and generation .

The analysis part performs all the processes related to the clauses entered and their analysis in away that guarantees obtaining their vocabulary items . In the dictionary part , used and modified by adding English vocabulary items for the meanings of verbs and their contexts in a way that assures obtaining their right references . Final , the generation part consists of a group of processes that organized the target sentences in the language of translation depending on the grammatical strucure that results from the source sentence.

The research also includes the use of semantic theories and methologies for the verbal Arabic sentences in an attempt to prove the competence of these theories and methodologies in describing the Arabic sentence semantically and according to context .

Keywords: Machine Translation;Transfer Approach;Arabic Language Processing.

1. Introduction

Machine Translation (MT) is the area of information technology and applied linguistics dealing with the translation of human languages such as Arabic and English . with globalization and expanding trade , demand for translation is set to grow. Computer technology has been applied to technical translation in order to improve one or both of the following factors [1] :1) speed 2) cost: computer aids in translation can reduce the cost per word of a translation . In addition , the use of MT can result in improvements in quality , particularly in the use of consistent terminology within a text or for a particular kind of client . As English is a universal language , most of the researches in Arabic MT are mainly concentrated on the translation between English And Arabic .This will help in simplifying the Arab communication with other countries . these systems are based mainly on the transfer model [3,4,5].

Rafea (1992) developed an English Arabic MT system , which translates a sentence from the domain of the political news of the middle East .Ibrahim (1991) discussed the problem of the English Arabic translation of the embedded idioms and proverb expressions in the English sentences . El-Desouki(1996) discussed the necessity of modular programming for English Arabic MT. Mokhtar (2000) developed an English Arabic MT system , which is applied to abstracts from the field of Artificial Intelligence.

On the contrary , little work has been done in developing Arabic – English MT systems.Al Barhamtoshy (1995) proposed a translation method for compound verbs . Shaalan (2000) described a tool for translation the Arabic interrogative sentence into English . Chalabi (2001) presented an Arabic – English MT engine that allows any Arabic user to search and navigate through the Internet using the Arabic language.

Athman (2003) developed an efficient chart parser that will be used for translation Arabic sentence.

The present work addresses the translate of a fairly complex Arabic verbal sentences into English .Which is an important task for automating the translation between Arabic and English sentences.

The next section outlines the overall architecture of the proposed Arabic- English machine translation system. The following sections describe the main components of the system . We also describe how we evaluated the correctness of our MT system . In a conclusion section , we discuss its applications on real verbal sentences of these from student – school domain and present some final remarks [6,7] .

2. Overall Structure of The System

There are three basic approaches being used for developing MT systems that differ in their complexity and sophistication. These approaches are : direct approach , transfer –

based approach and interlingual approach. There are many factors which make transfer an attractive design for MT .

1) many systems are bilingual , or their principles use is for translation in one direction between a limited number of language , 2) where full multilinguality is required , it is possible to have a hub language into and out of which all translation is done; and 3) portions of transfer modules can be shared when closely related languages are involved. For example, an English – portuguese module many share several transformations with an English – Spanish module. The architecture of the transfer Arabic to English MT system is given in fig. 1 with three main components : an analyzer component , a transfer component , and a generation component .

The programming language C++ is used for the time in computational linguistics because it is a modern and competent language in terms of its wide programming application to the construction of complex data structures , this language also has features object oriented languages which are lead to design the structure dictionary. The validity of the model has been proven by applying and discussing a number of examples.

3. Syntax Analysis

This stage can be find all the (words) or (terms) that constitute the source (VS). The development of the parser is a two step process . In the first step , we a acquire the rules that constitute a grammer for the Arabic (VS) that gives a precise account of what it is for a (VS) to be grammatically correct . The grammer covers the (VS) from the student and school domain . The grammer acquired from the analysis of 40 verbs.

Our analysis indicates the all contex the verb in the (VS). The second step is to implement the parser that assigns grammatical structure onto input (VS).

In order to implement the parser , it was needed to perform morphological analysis on the inflected Arabic words . The dictionary which is uses successfully implement the morphological analyzer.

The morphological analyzer returns to the number of noun . Entries of the word dictionary can be stems of verbs , noun. By another words the morphological analyzer can be disided if the (term) is true or not , this process is done by raised the Affixes , Suffixes and lookup in the dictionary.

4. Syntactic Transfer MT

Syntactic transfer system rely on mapping between the surface structure of sentences . The analysis tree of the source language in order to construct a target language analysis tree [8] . Top – down processing which one side of the tree transfer rules is matched a gainst the input structure . The transfer component has dictionary to perform Arabic to English translation . In our (VS) translation , the actual translation occurs in the transfer phase.

The following paragraphs describe the problems encountered while designing this phase . These problems are regards as peculiar to translation , since they arise from the divergences and mismatches between source and target VSs.

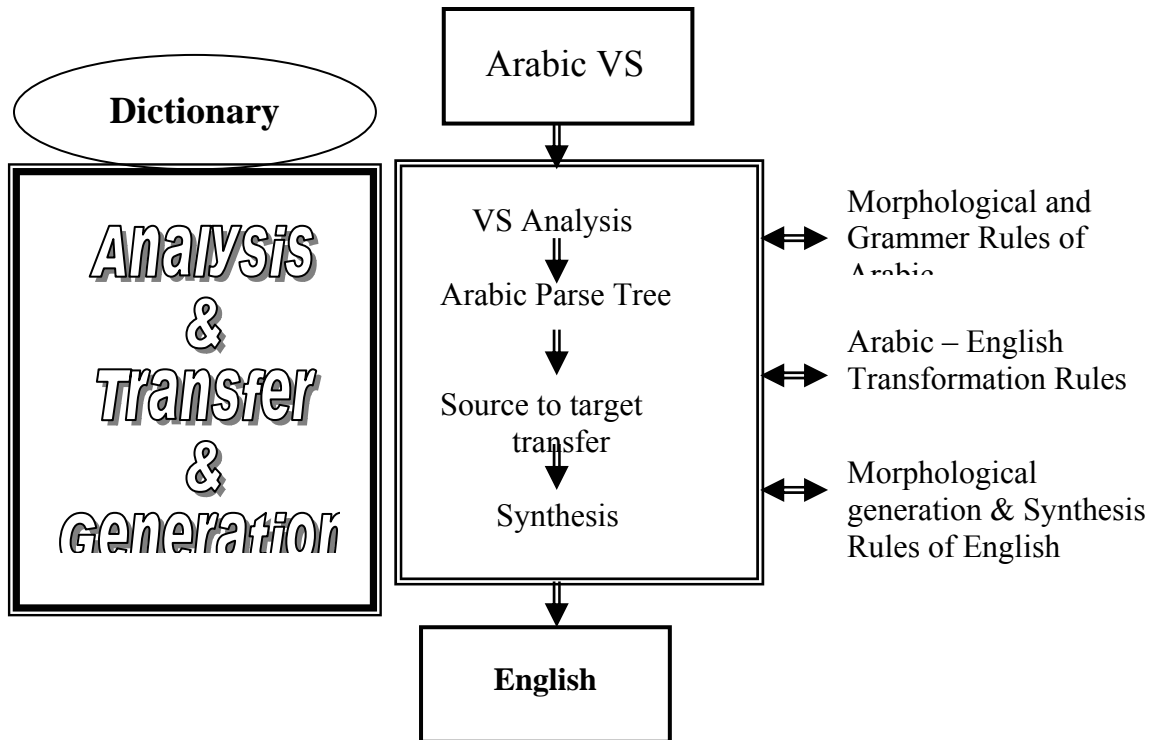


Figure 1. Overall structure of Arabic – English translator

5. Lexicon

We select the lexicon that represent the necessarily knowledge which are benefit in MT. fig.2 explain data structure representation. This data structure explain following [9 , 10] :-

Suppose the sentences is

كتب الطالب الدرس

The words are composed of this sentence is

كتب ال طالب ال درس

The (stem) is represent the entry of lexicon , our lexicon has two entry one for verb and another for name like this.

كتب ---> verb entry

طالب ---> name entry

درس ---> verb entry

name entry

we note the word (درس) has two entry verb and name . This type is upon on the contex sentences .

such as

درس الطالب الدرس

(Name) (verb)

كتب الطالب الدرس
Name Name Verb

The structure of lexicon is present solve for two problems Arabic language : 1) Syntactical Ambiguity , 2) Semantical Ambiguity.

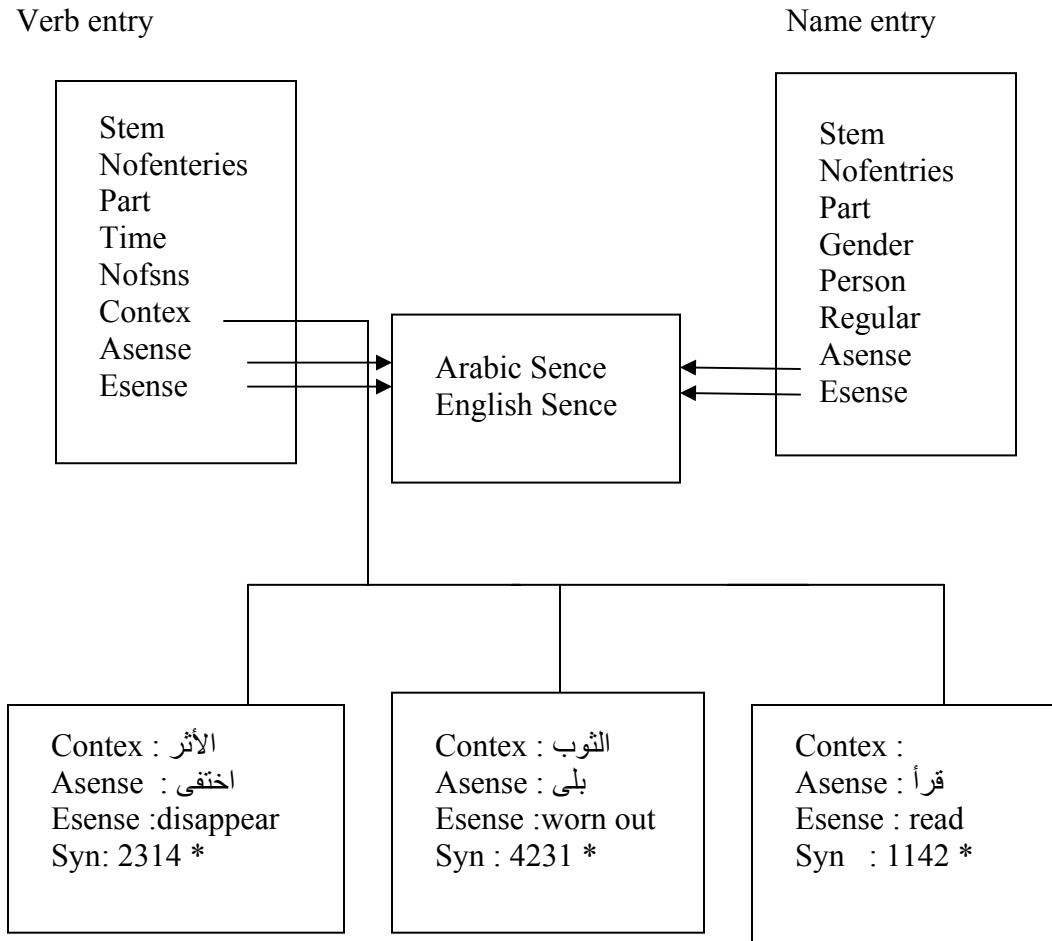


Figure 2. Word Structure Representation In the Lexicon

* This value find from the tables of Semantic Fields
The above structure for like this example:

درس الطالب الدرس
درس الأثر
درس الثوب

6. Translation Processing

after Analysis and parsing the verbal Arabic sentences lookup from the words, in the lexicon this processing came after morphological phase , if successful search go to generate phase , else call the morphological phase or exit from system in message “there is no word in lexicon” [11] .

7. Generate Phase

This phase is the last processing in the MT system , the work is rearrangement the words in the target sentence in terms at the grammar of English sentences , this phase is benefit from the input source Arabic sentences like this [12].

In Arabic : كتب الطالب الدرس
O S V

The form in English must be as

The Student wrote the lesson
SVO

And so on for all the types of sentences which are contains the preposition or adjective etc.

The arrangement of sentences in the English language is uses the grammar in this language and designed the processes uses finite state machine for matching the syntax grammar for the target sentences. Fig. 3 explain some examples of verb (كتب) .

Target sentence	Source sentence
The student wrote the lesson	كتب الطالب الدرس
The student wrote the lesson in the school	كتب الطالب الدرس في المدرسة
The hard student wrote the lesson	كتب الطالب المجد الدرس
The student wrote the lesson in the school	كتب الطالب في المدرسة الدرس
Error in source sentence	كتب الدرس
The hard student wrote the lesson	كتب الدرس الطالب المجد
The student wrote the lesson by pen	كتب الطالب الدرس بالقلم
The hard student wrote the lesson in the school	كتب الطالب المجد الدرس في المدرسة

Figure. 3 examples of verb (كتب)

8. Conclusion

This paper concentrated one the issues in the design and implementation MT system , which translates a verbal sentences Arabic into English . We showed that the MT approach is promising and can be used to automate the translation of the many verbal sentences in the student and school domain .

We have collected 50 verbs and their contex , in future works , the problems mentioned can be improved by storing in lexicon synonyms of words , implementing the

transfer that deals with the special case for mapping Arabic into English , and implementing the synthesis rules that introduces new words in the output translations for some special cases.

The structure of lexicon has present many facilities for backup upon word and Analysis the sentences also can be used in MT from Arabic into any other language by adding sense of that language.

Final there are comparison between results (system output) of MT and Al-kafi translator program pakage like fig. 4 which explain differencies between two systems , from this fig. we notice that , Al-kafi program depended on the default semantic not the other types semantics. Also al-kafi program not important in the multiples contex of the verb in the sentence , but our MT system concerned on this side, therefore , the efficiently results of proposed system more better than Al-kafi prgram as you like in fig. 4 .

Is the school the gold of the student	The student went to school	
Is its death , the gold of the student	The student died	
The student went away with the road sign	The student escorted the teacher	
The student went away	The student went	
The water carried in intoxicant off	The water mixed up with sucra	
The student ignored the school	The student left the school	
The student studied		
The cloth studied	The dress is worn out	
The student read the peace	The student greeted	
The student read	The first student read	

Figure. 4 The outputs for MT system and Al-kafi program

Refrences

1. A.Al Barhamtoshy , Arabic to English Translator of Compound Verbs , in Proceedings of the Annual Conference on Computer Science & Statistical ISSR, December 1995.
- 2.Dorr B.J.Garman J.,Weiberge A.,1995,” From Syntactic encodings to thematic roles: building lexical entries for interlingual :” university of mary land , the Nether lands.
- 3.Tahir Y.& Chenfour N.&Harit M.,2005, “Realization of amorphological analyzer for Arabic Language text”, Email :yousif.tahir@caramail.com.
- 4.Hutchins J.,2006 , “The State of Machine Translation in Europe”,university of east

- Anglia.
5. Elkateb S., and Block B., "English – Arabic Dictionary for translation ", Sabri – El-Kateb@student .
 6. Al – Shalabi R., 1996, "Design and implementation of an Arabic morphological system to support natural language processing", PH.D.thesis, computer science department , Illinois of technology .chicago TI, USA.
 7. J. Justeson and S. Katz, Technical Terminology :Some Linguistic Properties and an Algorithm for Identification in text , in Natural Language Engineering , Vol. 1, No.1, 9-27, 1995.
 8. D.J. Arnold , Sur la Conception du Transfert , La Traductique , P. Bouillon and A. Clas , (Eds.) , Les Presses de l'université de Montréal , Montréal, 1993, pp64-76. A version of this – in English "Transfer"- can be found in <http://clwww.essex.ac.uk/~doug/transfer.ps.gz>.
 9. D.J. Arnold , Lorna Balkan , Siety Meijer, R. Lee Humphreys and Louisa Sadler , Machine Translation : An Introductory Guide , Blackwells – NCC , London , 1994 , available at <http://clwww.essex.ac.uk/~doug/book/book.html>.
 10. K. Shaalan , M. Rafea and A. Rafea , KROL: A Knowledge Representation Object Language on Top of Prolog , Expert Systems with Applications: An International Journal , vol. 15, 33-46, 1998.
 11. E. Othman , K. Shaalan and A. Rafea , A Chart Parser for Analyzing Modern Standard Arabic Sentence , in Proceedings of the MT Summit IX Workshop on Machine Translation for Semitic Languages "Issues and Approaches , New Orleans , Louisiana , U.S.A., 2003.
 12. Abu Arafa A., 1995 , "A grammar for the Arabic Language suitable for machine parsing and automatic text generation " , PH.D. thesis Illinois of technology , Chicago , TI , USA.