Journal of
Applied Sciences

Translation of Classical Arabic Language to English

[1]Mohd Juzaiddin Ab Aziz and [2]Amna Mansur Hendr
[1]School of Computer Science, Faculty of Information Science and Technology,
University Kebangsaan Malaysia (The National University of Malaysia),
43600 Bangi, Selangor, Malaysia
[2]Department of Computer Science, Faculty of Qualifying Teachers,
Al-Mergheb University, Alkomes, Libya

Abstract: Machine Translation (MT) is the application of computers that translates texts from one natural language (source language) to another (target language). The past research dealt with problems mostly related to translating modern Arabic into English. This system is considered as the first of its kind to address the problem of translating classical Arabic into English where it involves cultures knowledge of the two languages. The work is a rule-based machine translation system and consists of three main modules, i.e., analysis, transfer and generation modules. In the transfer module phase, this research has developed and extracted the logical structure from Arabic and English to synchronize the sentences at lower level such as phrases. The generation module then combines the words and phrases to decide the appropriate meaning of them based on the situation of the sentences. A prototype was developed to prove the translation techniques that have been discovered. The performance of the system has been evaluated by comparing it with human translation. The accuracy of the results is 83.5%. These results proved the viability of this approach for Arabic-English machine translation.

Key words: Machine translation, classical Arabic language, transfer based approach, morphological, parsing, rule based approach

INTRODUCTION

Machine translation is an application of Natural Language Processing (NLP), commonly known as MT and it is a sub-field of computational linguistics that investigates the use of computer software to translate text or speech from one natural language (source language) to another (target language) (Abu-Shquier, 2009). MT is not simply replacing words with other words in target language but it needs complex system that contains linguistic rules such as morphology, semantics and syntax. Arabic natural language processing in general is still underdeveloped (Chaudhuri and Chaudhuri, 2006), because of the complexities of Arabic (Albared *et al.*, 2010). Moreover, tools used for other languages are not easily adaptable to Arabic due to the language complexity at both the morphological and syntactic levels (Abderrahim and Reguig, 2008). As Semitic language, Arabic has a rich derivational and inflectional morphology.

There has been much work on Arabic-English MT Most of this work are on translation of Modern Standard Arabic (MSA). However, the need for such a system is important in religious activities, Islamic education and poetry. It is also to overcome the language barriers when things related to Islam are discussed or debated.

The most important problem of the non-speakers in Arabic is how to understand everything about Islam and they do not understand classical Arabic and they needed to translate it to the language they understand. The translation of classical Arabic is a difficult task because translating the meaning from classical Arabic into English requires special traits in the translator and requires knowledge of two languages and cultures. In fact, there are several obstacles facing the translator when translation classical Arabic language to English such as-religious terms, the cultural differences between civilizations and religions, the lack of full efficiency and finding the appropriate meaning in the target language.

A researcher now has access to many systems, both commercial and research, of varying levels of performance (Pan *et al.*, 2012). All translation systems have the same goal: translate text in one language into text in a second language. Recently, most of the researches in machine translation focus on statistical approach and building a better probabilistic model. However, this approach is

---

Corresponding Author: Mohd Juzaiddin Ab Aziz, School of Computer Science, Faculty of Information Technology,
University Kebangsaan Malaysia (The National University of Malaysia), 43600 Bangi, Selangor, Malaysia

inadequate for translating classic Arabic to English due to the lack of bilingual parallel corpora. Moreover, Arabic presents an interesting problem for statistical models due to its rich and complex morphology (Frikha *et al.*, 2007; Saif and Aziz, 2011).

However, in the literature, several works have been proposed to address the translation from Arabic to other languages and vice versa. Most of these works are based on linguistic approach. Mohammed and Aziz (2011) applied transfer based approach to develop MT system to translate Arabic interrogative sentence to English in agriculture domain. In addition, the work by Shaalan *et al.* (2004) developed a machine translation that translates Arabic noun phrases to English. They also proposed a rule-based approach to solve these problems. Salem (2009) Also used transfer approach to translate English Noun Phrase (NP) into Arabic, The system is implemented in Prolog and the parser is written in DCG formalism. Shquier and Sembok (2008) develop rules based model to handle agreement and word-ordering problem in the context of English to Arabic MT. Salem and Nolan (2009) developed a system called UniArab to translate from Arabic to English using the Role and Reference Grammar Linguistic Model. They used the logical structure to represent Arabic sentences. They show how the characteristics of Arabic language influence the progress of MT tool. Mohammed and Aziz (2011) introduced a MT system which translates full text (abstract) written in English into Arabic.

CLASSICAL ARABIC LANGUAGE

Arabic language is one of the world major languages and it is usually divided into three distinct forms, namely, Classical Arabic (CA), Modern Standard Arabic (MSA) and Colloquial Arabic. Arabic is a language with a derivational and inflectional rich morphology (Abderrahim and Reguig, 2008; Albared *et al.*, 2011; Al-Salem and Aziz, 2011).

Classical Arabic is considered to be the most prestigious form of Arabic for all our modern writers in the sense that they try to follow the syntactic and the grammatical norms put down by classical grammarians such as Sibawaih and others (Al-Saidat and Al-Momani, 2010).

Many important documents were revealed in Classical Arabic which is the main reason why the language has preserved its purity throughout the centuries. Arabs consider Classical Arabic as an important part of their culture throughout Islamic history and it was the language used for all religious, cultural, administrative and scholarly purposes.

Table 1: Examples show difference of vocabularies

	Word	Example
Classical Arabic language	شتان	شتان بين الجنه والنار
Modern Standard Arabic language	اختلاف	هنالك فرق بين الجنة والنار
	فرق	هنالك اختلاف بين الجنة والنار
Classical Arabic language	صه	صه اذا لم تفهم
Modern Standard Arabic language	اسكت	اسكت اذا لم تفهم
Classical Arabic language	نضى	نضى الرجلُ السيفَ
Modern Standard Arabic language	انتزع	انتزع الرجلُ السيفَ
	استل	استل الرجلُ السيفَ
Classical Arabic language	مايبرح	مايبرح حتى يوذن له
Modern Standard Arabic language	مايترك	مايترك مكانه حتى يوذن له

Table 2: Examples show difference of Rhetoric

	Word	Example
Classical Arabic language	تكاكأتم إفرنقعوا	مالكم تكأكأتم عليّ كتكأكئكم على ذي جنةٍ .... افرنقعوا
Modern standard Arabic language	اجتماعم انصرفوا	مالكم اجتمعتم حولي كأنكم تجتمعون على مجنون...انصرفوا

The main difference between Modern Standard Arabic language and Classical Arabic language lies in the vocabulary. Table 1 examples describe some of these differences. For example, the word "شتان", in the first row, is not used nowadays but it is replaced with word "اختلاف" and also the other examples in same table show different between classical Arabic and modern standard Arabic language.

In Table 2, we can say that the Rhetoric is the art and study of the use of language with persuasive effect that is used frequently in literary techniques, poetry, etc. Through the meaning of the word "تكأكأتم" which is shown strong meaning in the classical Arabic language and evolved into the word "اجتمعتم" in MSA.

MATERIALS AND METHODS

The CArabicMT system is based on the transfer-based architecture with three major stages: An analysis stage, a transfer stage and a generation stage. This is done through some of translation process as according to Shirko *et al.* (2010).

The following summarized the main processes of the Rule based approach in the CArabicMT system:

- Input the source text in Arabic language (classical Arabic)
- Split the sentence (source text) into multiple sentences then split each sentence for many tokens (words)
- Morphological analysis: The Arabic morphological analyzer analyzes each Arabic word and extracts its features
- The syntactic parser builds a syntactic parse tree. The Parser starts to determine the structure of the sentence
- Lexical transfer module which looks for an equivalent English meaning of each word node in Arabic parse tree
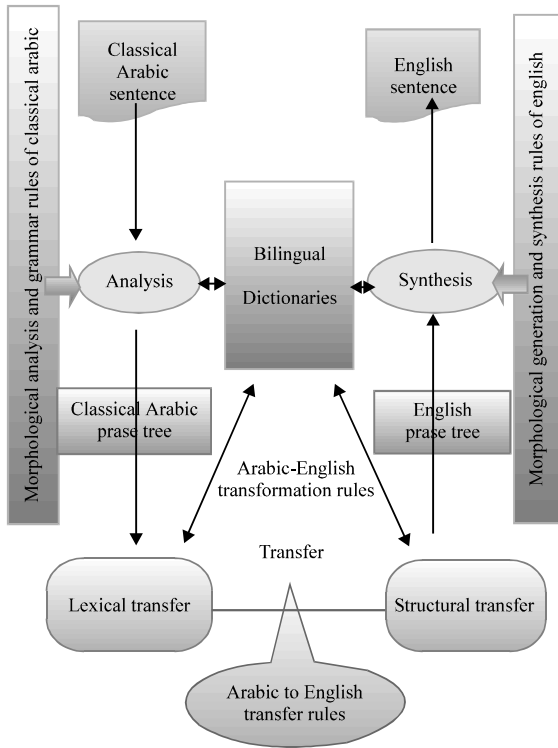
Fig. 1: The Architecture of classical Arabic-English MT

Fig. 2: Architecture of the parser

Table 3: Example of the morphological analyzer output

Suffix	Stem	Prefix	Arabic
	يزال	لا	لايزال
	ناس	ال	الناس
	خير	بـ	بخير
وا	عجل	ما	ما عجلوا
	فطر	ال	الفطر

- Structure transfer where the Arabic parse tree is transferred to the corresponding English parse tree
- After the transfer, the system will generate (synthesis) English sentence. The morphological generation module constructs the inflected English word; The syntactic generation module, in this step, traverses the English parse tree to generate the final structure of the English sentence

The architecture of CArabicMT system is given in Fig. 1.

Transfer-based machine translation: In this section, we will describe in more details the transfer approach and its three major stages: analysis stage, transfer stage and generation stage (Xiang *et al.*, 2011). Moreover, we will show, given examples from classic Arabic, how they process an Arabic sentence and translate it to its English equivalent.

Analysis stage: NLP systems and MT systems require identifying words in texts in order to determine their syntactic and semantic properties (Li *et al.*, 2009).

First of all, a classic Arabic sentence is passed through a tokenization module, in which the sentence is divided into tokens. By token, we mean the smallest syntactic unit; it can be a word, a part of a word
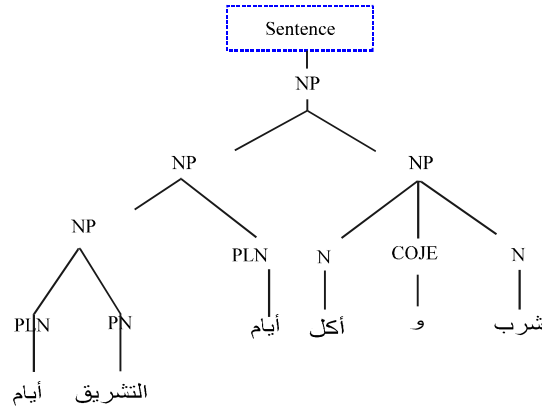
(Wang *et al.*, 2008). In analysis stage the information about source language only are needed.

Secondly, an Arabic morphological analyzer is used to extract stems of the Arabic words and to identify their morphological and syntactic properties such as gender, number, tense, etc.

Table 3 show an example of the morphological analysis output given the following:

" لا يزال الناس بخير ما عجلوا الفطر "

Syntactical analysis (Parsing) is the computer process of analyzing a sequence of tokens to determine its grammatical structure with respect to a given formal grammar; the parsing transforms input text into a data structure, usually a tree (Al-Saidat and Al-Momani, 2010). In the parsing, the classical Arabic sentence is represented in a tree of phrases, parse tree, where each phrase. Figure 2 shows the parse tree of an Arabic sentence.

Transfer stage: The transfer stage is a system of rules that relate words and structures in one language to words and structures of another language (Shirko *et al.*, 2010). Transfer starts with the output of the analysis phase and ends where the phase of generation starts (Shquier and Sembok, 2008). In the extent of this research; the translation actually occurs in the transfer phase. There are two types of transfer: First, lexical transfer, second, structural transfer.

The lexical transfer converts every word in the source sentence to corresponding target language representation (English word). The syntactic transfer convert the parse tree of the Arabic sentence to its equivalent in English. In addition, the bilingual dictionary is essentially in transfer method. Figure 3 illustrates an example of lexical transfer.

The structural transfer provides the rules for converting source language parser trees into equivalent target language trees (Trujillo, 1999). The transfer rules deals with the restructuring of the parse tree and reordering of words between the source and target language. Thus, this step provided the rules for convert Arabic parse tree into equivalent English tree. Figure 4 shows an example of the transfer step.

Generation: In this step, the parsing rules of the target Language is used to produce the target sentence. The generation stage is generally divided into two parts, syntactic generation and morphological generation. Syntactic generation: the English parse tree is traversed to produce the final structure of the English sentence.
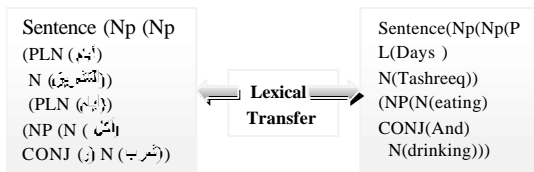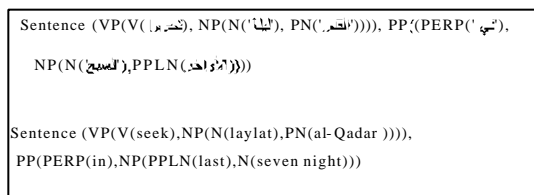

Fig. 3: Example of lexical transfer

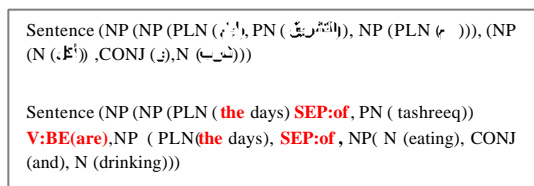Fig. 4: Example of the transfer process

Fig. 5: Example in generation stage

Morphological generation that generates inflected English word in its correct form based on a set of English grammar rules.

As an example, Fig. 5 shows how the generation rules added some English words (the, of, are) to produce the final structure of the English sentence:

Input   : أيام التشريق أيام أكل وشرب
Transfer : Days tashreeq day eating and drinking
Output   : The days of tashreeq are the days of eating and drinking

RESULTS AND DISCUSSION

In general, the rationale of this experiment is to investigate whether a machine translation system, Google and CArabicMT are sufficiently robust to be translated classical Arabic to English by comparing their output with the human translation. The test set is drawn randomly from a classic Arabic book "Sahih Muslim".

The result of evaluation of translation has performed by counting out of the total score of each system for all sentence, score is a number between 0 and 10 (Roshidul *et al.*, 2011; Jiang and Wei, 2012). Then, the score is divided it by the number of sentences multiplied by 10 to find their percentage out (Table 4).

The score is given by human expert in translation and it measures the differences between the human translation, Google and CArabicMT system. The range of scores is between 0 and 10 determines the correctness of the translation or matching translation according to the magnitude of error in the structure or the meaning of the generated English sentence. While 0 indicates an absolutely incorrect translation, 10 indicate absolutely correct (matched) translation.

The result was 49% for Google and 83.5% for CArabicMT. However, Table 5 represents all types of the

Table 4: Experiment results

Machine translation (MT)	Google	CArabicMT
Total score	98	167
Overall percentage	49%	83.5%

Table 5: Type of problem frequencies with classical Arabic to English MT

Problem No.	Type of problem	Total frequency	Google	CArabicMT
1	Synonyms of a word	38	36	2
2	Multiple word expression and Successive form	35	33	2
3	The preposition ambiguity	25	22	3
4	Order of the adjective	16	14	2
5	Addition and deletion	55	37	18
6	Lack of meaning or wrong meaning:	38	38	0
7	Word does not exist	40	40	0
8	The difference between the linguistic meaning and religious meaning	43	37	6
	Total frequencies of problem	290	257	33

Table 6: Shows part of the result produced by this experiment

Sentence (SL)	English (MT)	Translation	Original translation (TL)	Problem No.	Score
الصيام جنة	CArabicMT	The fasting is shield	Fasting is a shield		9
	Google	Fasting???		7	3
اذا رويته رافطروا لرويته فان غمي عليكم الشهر فعدوا للاثين سومرا	CarabicMT	Observe fast on sight it and break the fast on sight it and if it's concealed (cloudy) for you the month so count thirty	Observe fast on sighting it (the new moon) and break it on sighting it. But if (due to clouds) the actual position of the month is concealed from you, you should then count thirty (days)	5	9
	Google	Fast when you see it and stop fasting when you see ??? the Gumi Count your bad thirtieth month		678	4
أفضل الصيام بعد رمضان شهر الله المحرم وأفضل الصلاة بعد الفروضة صلاة الليل	CArabicMT	The best fast after the month of Ramadan is Allah's month Al-Muharram and the best prayer after the religious duty is night prayer	The most excellent fast after Ramadan is God's month. Al-Muharram and the most excellent prayer after what is prescribed is prayer during the night	58	8
	Google	The best fast after Ramadan is the month of Muharram and the best prayer after the obligatory night prayer		578	7
ما عجلوا الفطر الازال الناس بخير	CarabicMT	The people continue with prosper as long as they hasten the breaking of the fast	The people will continue to prosper as long as they hasten the breaking of the fast	35	8
لا شهرا عيد يتقصان رمضان وذو الحجة	Google	People still hurry up the fine-Fitr	The two months of 'Id, Ramadan and Dhu'l-Hijja (are not incomplete)	15678	3
	CarabicMT	Two months of EID are not incomplete the month of Ramadan and Dhu'l-Hijja			9
	Google	"Holiday not Enqassan month of Ramadan and is the argument		678215	2

Sentences have been tested by Google in December 2010

problem returned by each of the examined systems, Google and CArabicMT and their frequencies. If we examined the first row for the Synonyms of a word will find that this type of problem frequented 36 times with Google and only 2 times with our system. Therefore, as a total this type of problem frequented 38 times with two systems. Table 6 show some examples which explain in details the evolution process.

CONCLUSION

This study has been concentrated on issues in the design and implementation of a machine translation system which translates a sentence from classical Arabic language to English based on rule-based approach. There are several reasons that make transfer based is desired by MT community (Trujillo, 1999). These reasons showed that the transfer based approach is promising. So that we can improve the quality of machine translation output and increases its usefulness in classical Arabic. However, we have faced some difficulties when building the system, the first difficulty is; most of the existing Arabic database does not contain words in classical Arabic. However, due to a difference in the sentence structure between MSA and classical Arabic, it makes the rules which developed by them are not appropriate for the classical Arabic. Features classical Arabic is short sentence, word affecting and rhetoric miracle. Therefore, the translation classical Arabic directly is miracles even for specialists, also that Scholars of Arabic language stood to explain the vocabulary and interpretation of grammatical compositions and the development of its implications. Brevity sentences and abundance of meaning make the translation of classical Arabic more complexes. Therefore, the translation classical Arabic needs to have more sophisticated analysis and to go deep to the semantic and pragmatic levels to interpret its meaning. We think that will be one of the most challenges to the Arabic computational linguists in the future. However, the translation machines of the literature language such as poems and novels still in its fancy stage even in the good studied languages such as English. In future works, there are more activities are needed to enlarge and enrich our system, so to as better handle more complicated situations. All concerned with extending our paper work to be more powerful and applicable. In what follows, we presented some Proposals: Prepare and build a Private database of religious domain that containing the words written in classical Arabic and in the religious, development of this system to deal with formation. In addition translation of classical Arabic by Semantic translation that is the process of using semantic information to aid in the translation of classical Arabic sentence to create an equivalent meaning of this sentence. Here, we just reported our modest effort to open the research in this direction for further enhancement.

REFERENCES

Abderrahim, M.E.A. and F.B. Reguig, 2008. A morphological analyzer for vocalized or not vocalized Arabic language. J. Applied Sci., 8: 984-991.

Abu-Shquier, M.A., 2009. Word agreement and ordering in English-Arabic machine translation: A rule-based approach. Ph.D. Thesis, Universiti Kebangsaan Malaysia, Bangi, Malaysia.

Al-Saidat, E. and I. Al-Momani, 2010. Future markers in modern standard Arabic and Jordanian Arabic: A contrastive study. Eur. J. Soc. Sci., 12: 397-408.

Al-Salem, B. and M.J.A. Aziz, 2011. Statistical Bayesian learning for automatic Arabic text categorization. J. Comput. Sci., 7: 39-45.

Albared, M., N. Omar and M.J.A. Aziz, 2011. Developing a competitive HMM Arabic POS tagger using small training corpora. Intell. Inform. Database Syst. LNCS, 6591: 288-296.

Albared, M., N. Omar, M.J.A. Aziz and M.Z.A. Nazri, 2010. Automatic part of speech tagging for Arabic: An experiment using Bigram hidden Markov model. Rough Set Knowledge Technol. LNCS, 6401: 361-370.

Chaudhuri, M.A. and A.A. Chaudhuri, 2006. Transformation of spoken to written forms of (Natural) languages via spectral and pseudo-spectral methods. Trends Applied Sci. Res., 1: 350-361.

Frikha, M., Z. Ben Messaoud and A. Ben Hamida, 2007. Towards discriminative training estimators for HMM speech recognition system. J. Applied Sci., 7: 3891-3899.

Jiang, J.L. and W. Wei, 2012. Automated scoring research over 40 Years: Looking back and ahead. J. Artif. Intell., 5: 56-63.

Li, L., X. Wang and X. Wang, 2009. Weight-based feature selection for conditional maximum entropy models. Inform. Technol. J., 8: 764-769.

Mohammed, E. and M.J.A. Aziz, 2011. English to Arabic machine translation based on reordring algorithm. J. Comput. Sci., 7: 120-128.

Pan, X., H.B. Gu and Z.Q. Zhao, 2012. A novel feature selection framework in Chinese term definition extraction. Inform. Technol. J., 11: 148-153.

Roshidul, H., A.R.M. Shariff, B.S. Blackmore, I.B. Aris, A.R.B. Ramli and J. Hossen, 2011. Deploying natural language with topological relations for robotics behavior. J. Applied Sci., 11: 2993-3000.

Saif, A.M. and M.J.A. Aziz, 2011. An automatic collocation extraction from Arabic corpus. J. Comput. Sci., 7: 6-11.

Salem, Y. and B. Nolan, 2009. Designing an XML lexicon architecture for Arabic machine translation. Proceedings of the 2nd International Conference on Arabic Language Resources and Tools, April, 22-23, 2009, Cairo, Egypt, pp: 221-229.

Salem, Y., 2009. A generic framework for Arabic to English machine translation of simplex sentence using the role and reference grammar linguistic model. M.Sc. Thesis, Computing in the School of Information and Engineering, The Institute of Technology Blanchardstown, Dublin, Ireland.

Shaalan, K., A. Rafea, A.A. Moneim and H. Baraka, 2004. Machine translation of English noun phrases into Arabic. Int. J. Comput. Process. Languages, 17: 121-134.

Shirko, O., N. Omar, H. Arshad and M. Albared, 2010. Machine translation of noun phrases from Arabic to English using transfer-based approach. J. Comput. Sci., 6: 350-356.

Shquier, M.A. and T.M. Sembok, 2008. Word agreement and ordering in English-Arabic machine translation. Proceeding of the International Symposium on Information Technology, August 26-28, 2008, IEEE Xplore Press, Kuala Lumpur, pp: 1-10.

Trujillo, A., 1999. Translation Engines Techniques for Machine Translation. Springer-Verlag, Heidelberg, USA., ISBN: 1-85233-057-0111, pp: 220-222.

Wang, H., X. Sun, Y. Liu and Y. Liu, 2008. Natural language watermarking using Chinese syntactic transformations. Inform. Technol. J., 7: 904-910.

Xiang, L., X. Sun, Y. Liu and H. Yang, 2011. A secure steganographic method via multiple choice questions. Inform. Technol. J., 10: 992-1000.