

English to Arabic Machine Translation Based on Reordering Algorithm

Ehsan Abdulraheem Mohammed and Mohd. Juzaidin Ab Aziz
Department of Computer Science, Faculty of Information Technology,
University Kebangsaan Malaysia, Bangi, 43600, Selangor, Malaysia

Abstract: Problem statement: The purpose of Machine Translation (MT) is set accurate translation by implement a system to achieve similar result with human translates as in case the translating English to Arabic. The translation is a creative process which involves interpretation of the text that given by the translator. In addition, the translation depends on the purpose for which it meant according to context, as well as the text context includes different size and types of sentences. In addition, English words order in target language is not the same order as the Arabic source language, so the translation from English to Arabic definitely needs effort to synchronize words between English and Arabic that based on matching rules grammar between both languages. **Approach:** This study focuses on the existing Context Free Grammar (CFG) format as well as identifying the Part Of Speech (POS) for single words and reordering the CFG for reorder purpose of English structure to Arabic structure and validate the reorder CFG constructed by the algorithm. Reordering algorithm is a system for translating structured English sentence in text to structure Arabic sentence in text that runs compatibly with English/Arabic interface. English dictionary is used to translate single word consist of only word categories (POS) and Bi-lingual dictionary is used to translate single word consist of only word meaning relative to categories in approach format. **Results:** The domain area include twenty (20) abstracts containing ninety five (95) sentences have been tested in order to verify the authenticity of computer translation algorithm and the result were compared with human translation. The results obtain shows that the reordering rules is 81.855% accuracy on a translation from English Language to Arabic using an abstract from the European Psychiatry Journal. **Conclusion:** Based on the achieved results, we have managed to perform the syntactic reordering within an English sentence in text (abstract) to Arabic translation task by using Reordering Algorithm.

Key words: Machine translation, reordering, natural language processing, context free grammar, part of speech

INTRODUCTION

Machine Translation (MT) considers one of the oldest large-scale applications of computer science. It is also almost as old as the modern digital computer (Brown *et al.*, 1990). MT attempts to automate part or all of the process of translating from one human language to another such as English and Arabic (Arnold *et al.*, 1994; Zughoul and Abu-Alshaar, 2005) with or without human assistance. Depending on challenging, difficult and much of it is tedious and repetitive of work in this field, whereas requiring accuracy and consistency at the same time, as well as the huge request for translations in different domains. Therefore computer technology has been applied to technical translation to improve one or both of the following factors. First factor is speed that the translation by or with the aid of machines can be faster

than manual translation and second factor is cost that the computer aids in translation can reduce the cost per word of a translation (Shalan, 2000; Shalan *et al.*, 2010).

Arabic alphabet consists of 29 characters, where the shape of each character depends on its position within a word (Jannoud, 2007). The concept of machine translation English to Arabic received limited attention in recent years (Al-Taani and Hailat, 2005). Recently most of the researches in Arabic MT are mainly concentrated on the translation between English and Arabic that because English is a universal language and that will help in simplifying the Arab communication with other countries. There are some English to Arabic translation systems based mainly on the transfer model such as Alwafi and Al-Mutarjim, but they are at the beginners' stage when compared with the available translation. Al-Mutarjim is the commercial software that available from ATA software which translates

Corresponding Author: Ehsan Abdulraheem Mohammed, Department of Computer Science, Faculty of Information Technology, University Kebangsaan Malaysia, Bangi, 43600, Selangor, Malaysia

English text into Arabic. According to Shaalan *et al.*, (2004) study that mention some work in this area like "Ibrahim (1991) discussed the problem of the English-Arabic translation of the embedded proverb expressions and idioms in the English sentences, Rafea *et al.* (1992) developed an English-Arabic MT system, which translates a sentence from the domain of the political news of the Middle East, Maalej (1994) discussed the MT of English nominal compounds into Arabic based on their frequent occurrence in naming and referring in all text-types, Pease and Boushaba (1996) developed a system, which translates medical texts from English to Arabic and Mokhtar (2000) developed an English-Arabic MT system, which is applied to abstract from the field of Artificial Intelligence" (Shaalan *et al.*, 2004). There are systems used to translate Web pages from English to Arabic such as the system that uses commercial machine translation system to translate the textual part of a Web page from English to Arabic automatically, then displays a Web page containing the Arabic translation with all tags inserted in the right places and thus the layout and content of original (English) (Rached and Ahmed, 2001). This study present a system based on reordering words in each sentence in a text called Reordering Algorithm. In MT, global reordering is one of the major problems, since different languages have different word order requirement (Nguyen *et al.*, 2008). A word is "reordered "when its translation occupy different positions within the corresponding sentence (Zens *et al.*, 2004). English words order in source language is not the same order as the Arabic target language, so the translation from English to Arabic definitely needs effort to synchronize words between English and Arabic that based on matching between both language rules grammar. The Elming (2008) concludes from experimentation, 91% of the sentences contain reordering in order to synchronize with an average of 2.96 reordering per sentence. As a result for this study, to get accurate translation must consider that the sentence needs more than four times as many reordering per sentence. Current paper focus on a problem of get high quality translation from English into Arabic of full text that can be solved by first identify CFG format of English and match them with CFG format of Arabic then prepare translation rules set based on reordering CFG format and that rules related with both of the synchronization of the structure languages.

English and Arabic language structure: The structure of a sentence refers to how words in the sentence are related to each other. This structure indicates the way

that words are grouped together into phrases, what words modify what other words and what words are of central importance in the sentence. In addition, this structure may store other information about the particular sentence structure that may be needed for later processing and can identify the types of relationships that exist between phrases. A context-free grammar is a set of rules or production that expresses the ways of symbols of the language can be ordered and grouped together and a lexicon of words and symbols A CFG can be thought of in two ways: as a device for generating sentences and as a device for assigning a structure to a given sentence (Jurafsky and Martin, 2006).

When we mention Arabic throughout this study, we mean MSA, which is distinct from classical Arabic (Mahdi, 2005; Salem and Nolan, 2009). Arabic language have two main types of sentences are verbal sentences and equational or copula sentences. In verbal sentence, Arabic's relatively free word order allows the combinations of SVO, VSO, VOS and OVS for the elements of subject (S), verb (V) and object (O). Salem (2009) mention about report of Diab and Habash (2006) that refer to "the verb only appears in front of the subject 35% of the time. In another 35% of the cases, the subject appears in front of the verb, which is often due to topicalization. The remaining 30% of the time the subject does not appear since Arabic is a pro-drop language" (Salem, 2009). So the most common synchronize structures are SVO and VSO (Elming, 2008). Modern Standard Arabic (MSA) on the other hand has the basic word order VSO (Maamouri *et al.*, 2006), as well as Badr *et al.* (2009) refer to the VSO reordering frequent occurs more than SVO.

The declarative sentence is type of sentence in English that use in writing documents, essay and research therefore that kind of sentence is used in our domain. Per contra in English, the sentences have the basic word order SVO. English sentences consist of lexical items that in traditional grammar are called syntactic categories or Part Of Speech (POS). Part-Of-Speech tagging (POS tagging) is the process by which a specific tag is assigned to each word of a sentence to indicate the function of that word in the specific context (Elhadj, 2009). In Syntactic analysis stage start with detecting the basic grammatical categories of Verb (V), Noun (N), Proper (Prop), Pronoun (Proun), Abbreviation (Abbr), Adjective (Adj), Adverb (Adv), Preposition (Prep), Determiner (Det) Quantity (Quanti) and Conjunction (Conj). POS combine to form large units called phrases; these phrases could be noun (nominal), verbal, adjective, adverb and preposition. We need to identify the grammatical category phrase by

its head. A Noun Phrase (NP) identified when a phrase is headed by a noun; when a phrase is headed by a verb that called a Verb Phrase (VP). Those phrases combine to form larger units called clauses and sentence. Sentence have an internal structure that is POS as well as the phrases that contain them are hierarchically organized. There are rules of CFG that regulate and govern the internal structure of phrases and sentence that also called Phrase Structure Rules (PSRs). PSRs rules are to generate the sentence of the English. These PSRs are written using mathematical notation: $XP \rightarrow X$. This equation reads, a phrase of type X must have X as its head. The arrow means consists of, therefore if $X = N, V$, then $XP = NP, VP, AP$, respectively:

- $NP \rightarrow AP \ NP$
- $AP \rightarrow AdjP \ | \ AdvP$
- $AdjP \rightarrow (Adj) \ (\ AdjP)$
- $AdvP \rightarrow (Adv) \ (\ AdvP)$
- $NP \rightarrow to \ V$
- $NP \rightarrow Det \ NP$
- $NP \rightarrow Quanti \ NP$
- $NP1 \rightarrow NP \ Prep \ NP1$
- $NP1 \rightarrow NP$

There are major kinds of sentences and phrases in English of our domain that construct the structured English sentences and phrases. The major PSRs of English in our domain are:

- $S \rightarrow S; \ S$
- $S \rightarrow S, \ Conjunction \ S$
- $S \rightarrow S \ Conjunction \ S$
- $S \rightarrow Conjunction \ S, \ S$
- $S \rightarrow NP$
- $S \rightarrow (NP) \ (V) \ (S)$
- $NP \rightarrow Proun$
- $NP \rightarrow (N \ | \ Prop \ | \ Abbr) \ NP$

English to Arabic reordering rules: Here, we present the CFG rules used for reordering the English source to match with the CFG of the Arabic target. These rules are based on Arabic syntactic facts. Reordering the English can be done more reliably than other source languages, such as Arabic, Chinese and German since the state of the English parsers are considerably better than parsers of other language (Badr *et al.*, 2009). The following rules for reordering at the sentence level and the noun phrase level are applied to the English structure:

- NP: Reverse the order of nouns, adjectives and adverbs in the noun phrase are inverted. A result of

applying these rules, the phrase “very fast recovery” becomes “recovery fast very”

- PP: All prepositional phrase of the form $N1Prep \ N2Prep \ N2 \dots \ Prep \ Nn$ is transformed to $N1Prep \ N2Prep \ N2 \dots \ Prep \ Nn$ unless $N1$ of $N2$ of $N2 \dots$ of Nn is transformed to $N1 \ N2 \ N2 \dots \ Nn$. For instance, the phrase “geographical distribution of demand for psychiatric services” becomes “distribution geographical demand for services psychiatric”. As well as if phrase contains “of” and before the prepositional “of” is noun phrase and these NP contains sequence of nouns and before this NP the definite article “the” then delete also definite article “the” with prepositional “of” like the phrase “The contribution of computerized mapping techniques” becomes “contribution techniques mapping computerized”
- The: The definite article “the” is replicated before adjectives and this rule is applied NP in noun phrase after reverse. For example, “The effective clinical treatment” becomes “the treatment the clinical the effective”. The definite article “the” is also adding after prepositional and before NP in genitive like the phrase “The new thrust in community care” becomes “The new the thrust in care the community”
- VP: This rule transforms SVO sentences to VSO. All verbs are reordered on the condition that they have their own subject noun phase and are not in participle form since in these cases the Arabic subject occurs before the verb participle. The following Example illustrates all these cases: “the patient need a special treatment” becomes “need the patient treatment special”

MATERIALS AND METHODS

The system contains of two main phases. The first phase is the source language. In this phase divide the English sentences into parts until reach to words level as well as manipulates that English sentences generating suitable grammatical category. The second phase is the target language phrase. This phrase specifies one Arabic meaning for each word and aligns the target language words according to the target language rules.

The source language phase: This phase consists of six main steps:

- Divided the input text (abstract) into sentences: The output of this step is filling each record in Sentence_Arabic table with one by one sentence in

order based on source text (abstract) order sentence, but those sentences in Arabic structure.

Step 1: Import abstract into the system to process.

Step 1.1: Ask the user to insert input.

If the user chose an input from text file or insert text by manually into main form of the system Then

The system will move to step 2.

Otherwise, ERROR MSG will appear.

Step 2: Divided the input text into sentences that the input text will be broken down into sentences which ended with punctuation dot ‘.’.

Step 2.1: Divided the paragraph into sentences according to the punctuation dot ‘.’ and that sentences insert into sentences matrix.

Step 2.2: For each sentence in sentences matrix generated from step 2.1 Do

a. Sent that sentence to step 4.

b. Insert the output of step 4 as record into Sentence_Arabic table.

Step 3: Go to step 9.

Divided the sentence into subsentences: This step is important that for the structure of English. In this step, the sentence parsed into constituent’s conjunction and punctuation structure that will be needed to apply PSRs mentioned above. The system works only for structured English sentences only, which satisfied the PSRs. We have used top-down parse tree technique to check the internal structure of the English sentence. The input to this step is the sentence that import from previous step (from sentences matrix). The output of this step is a variable (sub_Arabic) that store all subsentences in sentence in order without any non-English character that means with target (Arabic) order structured. For understanding follow pseudo code for this step:

Step 4: Divid the sentence into subsentences, the sentence will be broken down into subsentences and that subsentences insert into subsentence matrix.

Step 4.1: Divid the sentence into subsentences according to find the punctuation or conjunction or second verb in sentence when this verb not prior by “to” then adds each subsentence into subsentences matrix.

Step 4.2: For each subsentence in subsentences matrix generated from step 4.1 Do

a. Sent that subsentence to step 5.

b. Insert the output of step 5 as element into sub_Arabic variable.

Divided the subsentence into phrases: In this step, the subsentence parsed into constituent’s noun, verb, adjective, adverb, structure that will be needed to apply PSRs mentioned above. The output of this step is correct grammatical category of each phrase in the subsentence and top-down parse tree for each subsentence structure. The input to this step is the subsentence that extracted from previous step (from subsentences matrix). The output of this step is a variable (phrase_Arabic) that store all phrases in subsentence in order without any non-English character that means with target (Arabic) order structured. For understanding follow pseudo code for this step:

Step 5: Extract the phrases based on previous study of English phrases of sentence on PSRs, The subsentence will be broken down into phrases and that phrases insert into phrases matrix.

Step 5.1: Extract the phrases based on previous study of English phrases of sentence on PSRs.

Step 5.1: For each phrases in phrases matrix Do

a. If it found in verb phrase (is, are, was, were) Then

Remove that word.

b. Sent that phrase to step 6.

c. Insert the output of step 6 as element into phrase_Arabic variable.

Divided the phrase into subphrases: In this step, the phrase parsed into constituent’s noun, adjective, adverb structure that will be needed to apply PSRs mentioned above of noun phrase . We also have some process before align these words in Arabic Language. For source language subphrase and other parts, the structure English sentence must add “the” in three situation. First, after preposition and before subphrase (noun phrase). Second, before sequence nouns which prior by “of”. Third, before subphrase (NP) when not prior by (the / a / an). It must remove “of” from phrases. The input to this step is the phrase that import from previous step (from phrases matrix). The output of this step is a variable (subphrase_Arabic) that store all subphrases in order without any non English character that means with target (Arabic) order structured. For understanding follow pseudo code for this step:

Step 6: Extract the subphrases through finding the noun phrase from the phrase, its will extract noun phrase as subphrases and add it to subphrase store.

Step 6.1: Extract subphrases from the phrase if found the noun phrase which contains nouns or adjectives or adverbs and that subphrases inserted into subphrase matrix.

Step 6.2:

- a. If the subject and /or object in SVO contain preposition and follow by noun phrase Then Add “the” in front the noun and adjective in noun phrase.
- b. If the subject and /or object in SVO contain “of” and the words in front of “of” are noun or sequence nouns that prior by “the” Then Remove “the” from phrase.
- c. If the subject and /or object in SVO contain “of” and the phrases in front and back of “of” is noun phrase Then Remove “of” from phrase.
- d. If the noun phrase in subject or object on SVO not prior by article “a/an” Then Add “the” in front of the nouns and adjectives in that noun phrase.
- e. Sent that phrase with its subphrases to step 7 and insert the output into subphrase_Arabic variable.

The target language phase:

Specify the Arabic meaning for each word: In this step, the phrase is ready to be translated into target language, word by word and in the same order in as the source phrase. We search the data based for the list of words that satisfy the query where the English word is the keyword with extract category (POS) for this word. The output of this step a list of Arabic words that gives the possible meanings for the corresponding English word. The input of this step is the subphrase_Arabic, phrase_Arabic variable and the outputs of this step are Arabic_words_phrase variable:

Step 7: Specify the Arabic meaning for each word

Step 7.1: a. Spilt the phrase with its subphrase into its contained English words in word [No] matrix, the order of word is very important.

b. The number of words in the sentence will assign to the variable Sentence_Length.

Step 7.2: For each single word [No] in the phrase Do

Step 7.2.1: Search for the word [No] in Bilingual dictionary:

If the word [No] found Then

Export the Arabic meaning into Arabic matrix and POS into POS matrix of that word at position [No].

Step 7.2.2: Move to the next word ([No+1]).

Step 7.3: a. Sent the Arabic metric, POS matrix to step 8.

b. Insert the output of step 8 into Arabic_words_phrase variable.

Align the target words according to target language

rules: Now we have raw material for a structure Arabic sentence, a set of lexical items not in the Arabic language correct order. We have some rules in the Arabic language to align these words. Arabic grammar rules list rules to contract the Arabic phrase. For the source language phrase, the structured English sentence must consists subject, verb and followed by an object, but the target language phrase, the structured Arabic sentence must consists verb, subject and followed by an object. The input to this step is the Arabic_words_phrase variable that generate in the previous step. This step is the reorder Arabic_words_phrase variable in the target language order:

Step 8: Aligning the target words according to the target language rules

- a. The verb placed in the left most of the sentence.
- b. If the subject and /or object in SVO contain nouns or adjectives or Adverbs both of them (subphrase) then swapping the order in reverses order.
- c. If the subject and /or object in SVO contain conjunction or quantity or pronoun or abbreviation Then Add conjunction or quantity or pronoun or abbreviation in the same order.
- d. If the subject and /or object in noun phrase in SVO contain the article “the /ال” Then Add “ال” in front each noun and adjective in that noun phrase.
- e. Insert the noun phrases that acts as the subject after the verb.
- f. Add the noun phrase that contains the object to the right most of phrases.

Presents the full abstract sentences in target

language (Arabic): The output of this step is the final sentences which system generates in the target language structure. This output presents in interface after import

them from Sentence_Arabic table in order. Finally, delete all records from Sentence_Arabic table that to free more space for next abstract process:

- Step 9: Presents the full abstract sentences in target language (Arabic)
- Export records in order from Sentence_Arabic table and generate target text.
 - Delete all records from Sentence_Arabic table.

Experimental: In general, the aim of this experiment is show the process of Reordering Algorithm based on

methodology. First insert the input text (abstract) to the system then apply step 1, 2, 3, 4, 5, 6, 7 and 8 (divided the input text (abstract) into sentences, divided the sentence into subsentences, divided the subsentence into phrases divided the phrase into subphrases, specify the Arabic meaning for each word and align the target words according to the target language rules. The following table summarized this process (Table 1). See the article (the) in table means this additional into in table based English to Arabic reordering rules that presents in previous section.

In step 9, from last column in Table 1, presents the abstract sentences words in target language structure (Arabic) (Table 2).

Table 1: Abstract in Arabic with process of reordering algorithm

Abstract in English	Sentence	Subsentence	Phrase	Subphrase	Word	Words-Arabic before reordering	Words_Arabic after reordering	
Spatial (geographic) information systems have an important role in psychiatric epidemiology. Drawing on a New-Zealand case study, this paper will evaluate the contribution of computerised mapping techniques in projecting the geographical distribution of demand for psychiatric services. A review the literature, is followed by a discussion of the methodology used in this research for forecasting demand for psychiatric care. The paper will also highlight the importance of census data for the planning of psychiatric services at regional and district levels.	Sentence1	subsentence1	phrase1		(the)	ال	تملك	
							ال	
				subphrase1	spatial	مكاني	ال	نظم
				subphrase1	(geographic)	(جغرافي)	معلومات	ال
				subphrase1	information	معلومات	(جغرافي)	ال
						(the)	ال	ال
				phrase2	-----	systems	نظم	مكاني
				phrase3	-----	have	تملك	-----
				subphrase1	important	مهم	دور	-----
						role	دور	مهم
						in	في	في
						(the)	ال	ال
				subphrase2	psychiatric	طب النفسي	علم الاوبئة	ال
				subphrase2	epidemiology	علم الاوبئة	طب النفسي	ال
		sentence2	subsentence1	phrase1		Drawing	بالاعتماد	بالاعتماد
				-----	on	على	على	على
				-----	a	أ	-----	-----
					subphrase2	New-Zealand	نيوزلند	دراسة
						case	حالة	حالة
						study	دراسة	نيوزلندا
		subsentence2	phrase1		this	هذه	سوف	
					(the)	ال	تقيم	
				subphrase1	paper	ورقة	هذه	
			phrase2		will	سوف	ال	
					evaluate	تقيم	ورقه	
			phrase3		the	-----	-----	
				subphrase1	contribution	مساهمة	مساهمة	
					of	-----	-----	
					(the)	ال	ال	
			subphrase2	computerised	(the)	حاسوبية	تقنيات	
						ال	ال	
				subphrase2	mapping	تخطيطه	تخطيطه	
					(the)	ال	ال	
				subphrase2	techniques	تقنيات	حاسوبية	
					in	في	في	
					(the)	ال	ال	
				subphrase3	projecting	عرض	عرض	
					the	ال	ال	
				subphrase4	geographical	جغرافي	توزيع	
					(the)	ال	ال	

Table 1: Continue

		subphrase4	distribution	توزيع	جغرافي
		-----	of	-----	-----
			(the)	ال	ال
		subphrase5	demand	مطلوب	مطلوب
		-----	for	ل	ل
			(the)	ال	ال
		subphrase6	psychiatric	طب النفسي	خدمات
		-----	(the)	ال	ال
		subphrase6	services	خدمات	طب النفسي
		-----	A	أ	-----
sentence3	subsentence1	phrase1	review	مراجعة	مراجعة
		-----	the	ال	ال
		subphrase2	literature	مطبوعات	مطبوعات
	subsentence2	phrase1	is	هي	-----
		-----	followed	تبع	تبع
			by	بواسطة	بواسطة
	phrase2	-----	a	أ	-----
		subphrase1	discussion	مناقشة	مناقشة
		-----	of	-----	-----
			the	ال	ال
		subphrase2	methodology	منهجية	منهجية
		-----	(the)	ال	ال
		subphrase2	used	مستخدمة	مستخدمة
		-----	in	في	في
			this	هذه	هذه
		-----	the	ال	ال
		subphrase3	research	بحوث	بحوث
		-----	for	ل	ل
			(the)	ال	ال
		subphrase4	forecasting	تنبؤ	تنبؤ
		-----	(the)	ال	ال
		subphrase4	demand	مطلوب	مطلوب
		-----	for	ل	ل
			(the)	ال	ال
		subphrase5	psychiatric	طب النفسي	رعاية
		-----	(the)	ال	ال
		subphrase5	care	رعاية	طب النفسي
	subsentence1	phrase1	The	ال	سوف
		-----	paper	ورقة	ال
	phrase2	-----	will	سوف	ورقة
	phrase3	-----	also	أيضاً	أيضاً
		subphrase1	highlight	تبرز	تبرز
		-----	the	-----	-----
		subphrase2	importance	أهمية	أهمية
		-----	of	-----	-----
			(the)	ال	ال
	subphrase3	census	إحصاء السكاني	بيانات	ال
		-----	(the)	ال	ال
		subphrase3	data	بيانات	إحصاء السكاني
		-----	for	ل	ل
			the	-----	-----
		subphrase4	planning	تخطيط	تخطيط
		-----	of	-----	-----
			(the)	ال	ال
		subphrase5	psychiatric	طب النفسي	خدمات
		-----	(the)	ال	ال
		subphrase5	services	خدمات	طب النفسي
		-----	at	على	على
			(the)	ال	ال
		subphrase6	regional	اقليمي	مستويين
		-----	and	و	ال
			(the)	ال	اقليمي
		subphrase7	district	محلي	و
		-----	(the)	ال	ال
		subphrase8	levels	مستويين	محلي

Table 2: Abstract in Arabic

Abstract in Arabic
تملك النظم المعلومات (الجغرافي) المكاني دور مهم في علم الأوبئة الطب النفسي. بالاعتماد على دراسة حالة نيوزلند، سوف تقم هذه الورقة مساهمه التقنيات التخطيطية الحاسوبية في العرض التوزيع الجغرافي المطلوب ل الخدمات الطب النفسي. مراجعه المطبوعات، تبع بواسطة مناقشة المنهجية المستخدمة في هذه البحوث ل التنبؤ المطلوب ل الرعاية الطب النفسي. سوف الورقة أيضاً تبرز أهمية البيانات الإحصاء السكان ل تخطيط الخدمات الطب النفسي على المستويين الاقليمي و المحلي.

RESULTS AND DISCUSSION

The domain area is abstracts from European Psychiatry Journal include twenty (20) abstracts containing ninety five (95) sentences have been tested in order to verify the authenticity of computer translation algorithm and the result were compared with human translation. English dictionary is used to translate single word consist of only word categories (POS) and Bi-lingual dictionary is used to translate single word consist of only word meaning relative to categories in approach format. A set of syntax-based reordering rules applied in several steps of the approach. When evaluating the system, a percentage of approximately 81.855% of the sample abstracts was translated onto Arabic and give correct translation.

CONCLUSION

This study has concentrated on the issues in the implements MT system, which translates full text (abstract) in English into Arabic based on Reordering Algorithm. We showed that the Reordering Algorithm is promising and used to automate the translation of abstracts of European Psychiatry Journal and that system get 81.855% of correct translation. This study also presents CFG format that most used in English and the linguistically motivated rules that reorder English to look like Arabic. In future works, solve the ambiguity of the meaning by create specialized lexicons, study the rest of CFG format which used in our domain that because this study focused on the most CFG format that used, but not all. These improvements will raise the correctness of the translations from 81.855-100%.

REFERENCES

Al-Taani, A.T. and E.M. Hailat, 2005. A direct english-arabic machine translation system. Inform. Technol. J., 4: 256-261. ISSN: 256-261. 1812-5638. DOI: 10.3923/itj.2005.256.261

Arnold, D.J., L. Balkan, R.L. Humphreys, S. Meijer and L. Sadler, 1994. Introduction. In: Machine Translation: An Introductory Guide, 1st Ed., Blackwells Publishers, London, pp: 200. ISBN: 1855542-17x

Badr, I., R. Zbib and J. Glass, 2009. Syntactic phrase reordering for english-to-arabic statistical machine translation. Proceeding of the Paper Read at Proceedings of the 12th Conference of the European Chapter of the ACL, Apr. 3, at Athens, Greece, pp: 86-93. DOI: 10.1.1.154.4570

Brown, P.F., J. Cocke, S.A.D. Pietra, V.J.D. Pietra and F. Jelinek *et al.*, 1990. A statistical approach to machine translation. Computational Linguistics, 16: 79-85. doi:10.3115/991365.991407

Elhadj, Y.O.M., 2009. Statistical part-of-speech tagger for traditional Arabic texts. J. Comput. Sci., 5: 794-800. ISSN: 1549-3636. DOI: 10.3844/jcssp.2009.794.800

Elming, J., 2008. Syntactic reordering in statistical machine translation. PhD Thesis. Copenhagen business school Denmark. http://openarchive.cbs.dk/bitstream/handle/10398/7922/jakob_elming.pdf?sequence=1

Jannoud, I.A., 2007. Automatic Arabic hand written text recognition system. Am. J. Applied Sci., 4: 857-864. DOI: 10.3844/2007.857.864.

Jurafsky, D. and J.H. Martin, 2008. Formal Grammars of English. In: Speech and Language Processing: An Introduction to Natural Language Processing Computational Linguistics and Speech Recognition, Calvino, I. G. Friedlander, J. Baldwin G.B. Trudeau and K. Kesey, (Eds.). Prentice Hall, USA., pp: 385-424.

Maamouri, M., A. Bies and S. Kulick, 2006. Diacritization: A challenge to Arabic treebank annotation and parsing. Paper read at The Challenge of Arabic for NLP/MT. International Conference at the British Computer Society, Oct. 23-24, SIG, London, UK., pp: 35-47.

Mahdi, A., 2005. Using Arabic: A Guide to Contemporary Usage. 1st Edn., Cambridge University Press, Cambridge, pp: 356. ISBN: 0 521 648327

Rached, N.Z. and A.G. Ahmed, 2001. An automatic English-Arabic HTML page translation system. J. Network Comput. Appl., 24: 333-357. DOI: 10.1006/jnca.2001.0135

Salem, Y., 2009. A generic framework for Arabic to English machine translation of simplex sentences using the Role and Reference Grammar linguistic model. M.Sc. Thesis. Computing in the School of Informatics and Engineering, The Institute of Technology Blanchardstown. http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg/Yasser_Salem_MSc_thesis.pdf

Salem, Y. and B. Nolan, 2009. Designing an XML Lexicon Architecture for Arabic Machine Translation Based on Role and Reference Grammar. Proceedings of the 2nd International Conference on Arabic Language Resources and Tools, April, 22-23, Cairo, Egypt, pp: 221-229.

- Shaalán, K., 2000. Machine translation of Arabic interrogative sentence into English. Proceedings of the 8th International Conference on Artificial Intelligence Applications, Dec. 3-6, Egyptian Computer Society, Egypt, pp: 473-483. DOI: 10.1.1.113.2443
- Shaalán, K., A. Hendam and A. Rafea, 2010. An English-Arabic bi-directional machine translation tool in the agriculture domain: A rule-based transfer approach for translating expert systems. IFIP Adv. Inform. Commun. Technolo., 340: 281-290. DOI: 10.1007/978-3-642-16327-2_34
- Shaalán, K., A. Rafea, A.A. Moneim and H. Baraka, 2004. Machine translation of English noun phrases into Arabic. Int. J. Comput. Process. Oriental Languages, 17: 121-134. DOI: 10.1142/S021942790400105X
- Nguyen, V.V., T.P. Nguyen, A. Shimazu and M.L. Nguyen, 2008. Reordering pPhrase-based machine translation over chunks. Proceeding of the 2008 IEEE International Conference on Research, Innovation and Vision for the Future in Computing and Communication Technologies, Jul. 13-17, IEEE Xplore Press, Ho Chi Minh City, pp: 114-119. DOI: 10.1109/RIVF.2008.4586342
- Zens, R., H. Ney, T. Watanabe and E. Sumita, 2004. Reordering constraints for phrase-based statistical machine translation. Proceedings of the 20th International Conference on Computational Linguistics, Aug. 23-27, Stroudsburg, PA, USA., pp: 219-225. DOI: 10.3115/1220355.1220385
- Zughoul, M.R. and A.M.I. Abu-Alshaar, 2005. English/Arabic/English Machine Translation: A historical perspective. J. Traducteurs Trans., 50: 1022-1041. <http://www.erudit.org/revue/meta/2005/v50/n3/011612ar.pdf>