A Hybrid Rules and Statistical Method for Arabic to English Machine Translation

Conference Paper · May 2019

DOI: 10.1109/CAIS.2019.8769545

3 authors:

Arwa Alqudsi
Universiti Kebangsaan Malaysia
16 PUBLICATIONS **101** CITATIONS
SEE PROFILE

Nazlia Omar
Universiti Kebangsaan Malaysia
90 PUBLICATIONS **1,033** CITATIONS
SEE PROFILE

Khalid Shaker
University of Anbar
34 PUBLICATIONS **262** CITATIONS
SEE PROFILE

A Hybrid Rules and Statistical Method for Arabic to English Machine Translation

Arwa Alqudsi/ Nazlia Omar
Knowledge Technology Research Group (KT), Center of Artificial Intelligence Technology (CAIT)
University Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia
e-mail: alqudsi@siswa.ukm.edu.my, nazlia@ukm.edu.my,

Khalid Shaker
College of Computer Science and IT, University of Anbar
Ramadi, Anbar,,Iraq
e-mail: khalidalhity@gmail.com

***Abstract*—Arabic is one of the six major world languages. It originated in the area currently known as the Arabian Peninsula. Arabic is the joint official language in Middle Eastern and African states. Large communities of Arabic speakers have existed outside of the Middle East since the end of the last century, particularly in the United States and Europe. So finding a quick and efficient Arabic machine translator has become an urgent necessity, due to the differences between the languages spoken in the world's communities and the vast development that has occurred worldwide. Arabic combines many of the significant challenges of other languages like word order and ambiguity. The word ordering problem because of Arabic has four sentence structures which allow different word orders. Ambiguity in the Arabic language is a notorious problem because of the richness and complexity of Arabic morphology. The core problems in machine translation are reordering the words and estimating the right word translation among many options in the lexicon. The Rule-Based Machine translation (RBMT) approach is the way to reorder words, and the statistical approach, such as Expectation Maximisation (EM), is the way to select right word translations and count word frequencies. Combining RBMT with EM plays an impotent role in generating a good-quality MT. This paper presents a combination of the rule-based machine translation (RBMT) approach with the Expectation Maximisation (EM) algorithm. These two techniques have been applied successfully to word ordering and ambiguity problems in Arabic-to-English machine translation.**

Keywords- Hybrid Approach; Expectation Maximisation; Arabic machine translation problems

I. INTRODUCTION

The Arabic language gained attention within the natural language processing (NLP) group because of its political significance and the language variations between it and other languages. The language features, especially the complicated morphology present endless difficulties for Arabic language scientists. Hybrid machine translation combines the primary of existing methods, including the rule-based method, statistical-based method, and example-based method, all which comprise the deficiencies of an individual machine translation strategy. RBMT is characterised by the explicit use and accurate manual creation of linguistically informed rules and representations by representing every piece of the input. However, this approach is limited by lexical selection in transfer and analysis failure sentences. The SMT method is more robust and provides fluent translation given the better lexical selection and use of language models. However, SMT encounters difficulties in dealing with linguistic knowledge, such as syntactic functions and word order, which lead to accuracy loss. Hybrid MT can be constructed based on RBMT. To overcome this, a hybrid system builds on statistical approaches by adding linguistics rules ([1]and [2]). This paper presents a combination of the rule-based machine translation (RBMT) approach with the Expectation Maximisation (EM) algorithm. These two techniques have been applied successfully to word ordering and ambiguity problems in Arabic-to-English machine translation.

II. RELATED WORK

There is comprehensive consensus in the literature that it is possible to produce hybrid translations from different systems perform in an improvement over the individual baseline systems ([1] and [3]). Hybrid techniques are used to incorporate advanced level subjective format guidelines to arrive at the final translation. Multiple techniques have been researched by analysts, but without any real success, as it was difficult to combine the fundamentally different techniques [7]. Information on new methods, i.e. how words should be converted, phrases and patterns, understanding of how syntax-based translation guidelines should be applied, and data of how syntactically centred target components have been developed [12]. Previous studies ([1]; [10]; [11]) reported that the result of combining rules with statistical methods could perform better translation than any other system. Such a combination can help produce a good-quality translation by addressing a wider range of explicit knowledge and syntactic constructions. [1] produced an organized category depending on the fact that mixtures of

978-1-7281-0108-8/19/$31.00 ©2019 IEEE

machine translation techniques are guided by a primary program which can be either concept or corpus-based. The category covered the majority of analysis on HMT. The authors reported that hybridisation of rule-based and corpus-based machine translation can be classified into those guided by either rule-based or guided by corpus-based machine translation. The former merge data information into a rule-based architecture which merges linguistic rules into a corpus-based architecture. [9] present techniques for integrating information from a rule-based machine translation system into a statistical machine translation to translate from Arabic to English. The techniques grouped into morphological, lexical, and system levels. The authors show how the rule-based process can be used to improve statistical machine translation. Morphologically rich languages pose a challenge for statistical machine translation (SMT) [6]. This challenge is magnified when translating to a morphologically rich language [8]. In machine translation from Arabic into English, syntactic, semantic, and Lexical problems arise and that poses many challenges, rich vocabulary, handling a complex and morphology [21].

[20] introduce a very rich Arabic morphological analyzer and generator called CALIMAStar, which provides form-based morphological features. The database used has six tables. Three are lexicon tables for stems, prefixes, and suffixes. The lexicon table has three columns: a lookup form, a list of feature-value pairs, and a compatibility category to control behavior and agreement.

[19] improved statistical machine translation using rule-based pre-processing and then coupling an Arabic-French statistical machine translation system using the Moses decoder with additional morphological rules that reduce Arabic morphology and makes it closer to French language morphology. [18] developed a hybrid machine translation system based on Arabic-to-English rule-based and example-based machine translation approaches. The system uses rule-based to induce the Arabic-to-English transfer rules from a set of example pairs. [19] improved statistical machine translation using rule-based pre-processing and then coupling an Arabic-French statistical machine translation system using the Moses decoder with additional morphological rules that reduce Arabic morphology and makes it closer to French language morphology. [18] a hybrid machine translation system based on Arabic-to-English rule-based and example-based machine translation approaches. The system uses rule-based to induce the Arabic-to-English transfer rules from a set of example pairs.

III. HYBRIDISATION RULE AND EXPECTATION MAXIMISATION APPROACH

The rule-based approach is the way to reorder words, and the statistical approach, such as EM, is the way to select right word translations and count word frequencies. Combining RBMT with EM is an improved solution to generate good-quality MT [1]. In this study, only the Maximisation part (M-part) of the EM algorithm is combined with the rule-based approach. The E-part is replaced with the rule-based approach because the task of this part is to address the word-ordering problem. This part fails to solve the problem completely because of the complexity of the Arabic language. The Expectation step in MT often fails because the translation provides different alignments for target language distribution that cannot be sampled with simpler change ([14] and [15]). By contrast, the rule-based approach can handle the word ordering problem in Arabic-to-English MT. In hybridisation, translations are performed using the RBMT system. SMT is then used to adjust the output from RBMT using the parallel corpus. Combinations of RBMT and SMT are highly competitive in MT quality. The output tends to be grammatical and the main effect of the combination is an increase in lexical selection quality, which strengthens one of the weak points of pure RBMT systems.

A. RBMT Component

The rule-based approach originates from the transfer system MT, where it in similarity uses rules ([4] and [5]), and is based on the linguistic information on the source and target languages that are mostly extracted from dictionaries. Given that rule-based knowledge exists in the structures of rules, system expansion becomes easy when the number of rules increases. High-quality translation can be produced when a large number of rules are used and deep-level linguistic analysis is conducted ([5]; [10]; [13]). The translation process of the RBMT approach consists of six main phases. This approach performs grammar rule matching for the translation. A significant difference exists between the Arabic and English grammar rules. Therefore, this approach matches these gaps and translates sentences. A considerable number of grammar rules are needed to perform the translation process and to enhance the accuracy of the translation output by adding the rules to the system. For further explanation, the example of the Arabic sentence "**المعلم يُدرس الطالب**". The manual translation of this Arabic sentence is "the teacher is teaching the student." The translation of this sentence passes the six main phases.

STEP 1: Tokenization. This is the first step in the rule-based part of the proposed hybrid approach. In tokenization, the input Arabic sentence is divided into a number of tokens, as shown in Figure 1. The task of the tokenizer is to divide the input text into tokens to define the demarcating numbers, words, clitics, and abbreviations. The tokened text is fed into a morphological transducer for further processing.

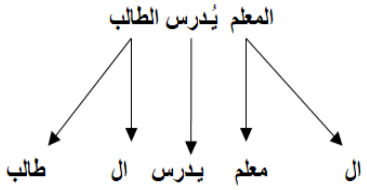
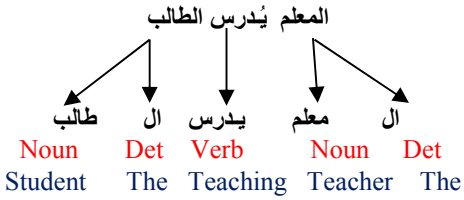


Figure 1 Tokenization process

STEP 2: Morphological analysis. In this step, the morphological analyser provides morpho-syntactic information on every token obtained from the lexicon. From the following example in Figure 2, the morphological analyser identifies each Arabic word in the sentence to determine their syntactic properties. The system depends on the morphological part to analyse words and understand the relationship among the different forms of one word and to determine which form can be taken.



Lexicon

Arabic word	Syntactic information	English
ال	Det	The
معلم	Noun	Teacher
يدرس	Verb	Teach
ال	Det	The
طالب	Noun	Student

Figure 2 Morphological analyzer for Arabic sentence

STEP 3: Parsing. After determining the syntactic properties of the Arabic words, the syntactic parser builds a syntactic relevant tree that represents the relationships among the words in the phrase. Parsing is a major aspect of MT systems, and it involves dividing a sentence into smaller sets depending on their syntactic functions in the sentence. According to Figure 3, the Arabic sentence is divided into two phrases, i.e., verb phrase (VP) and noun phrase (NP), to represent the phrase structure tree for the Arabic sentence " المعلم يُدرس الطالب." Its English equivalent is "the teacher is teaching the student." The rule-based parser maps the Arabic rules into English rules.

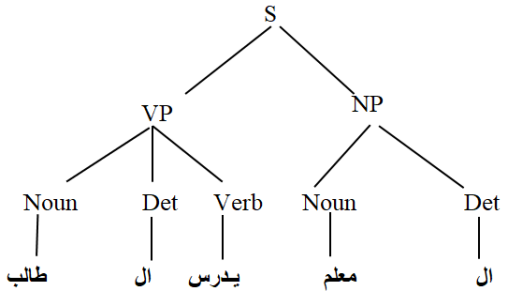


Figure 3 Parse tree for the Arabic sentence

STEP 4: Lexical dataset. The lexicon offers definite information on each individual lexical item word or phrase in the vocabulary of the language concerned. The grammatical rules identify probable grammatical structures in a language, and the lexicon declares which word can appear and in which construction.

STEP 5: Structure transfer. This step maps the Arabic dependency tree to the equivalent English syntactic structure. This step also provides the rules for mapping the parse tree of the source language into an equivalent parse tree in the target language, as shown in Figure 4. The Arabic lexical item is replaced with English lexical item. Transfer is the link between the two steps. In the transfer step, all features of translational differences (whether lexical or structural features) between the source and target languages are determined. Transfer begins with the output of the analysis phase and ends where the generation phase starts.

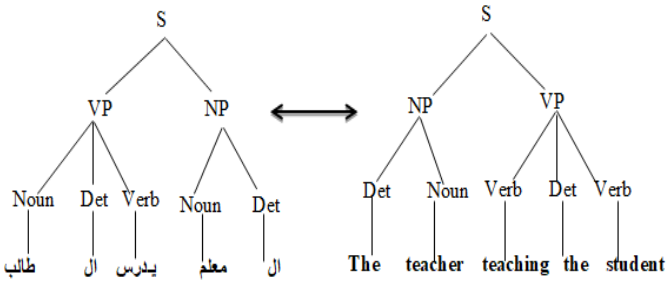


Figure 4 Arabic parse tree and its equivalent in English

STEP 6: Generation. In this step, all the lexical items and structures produced from the transfer step serve as input to produce well-formed and grammatically acceptable sentences in the target language. The generation process includes only target language information and works independently of the source language. In the generation, the sentence is generated by using a parser of the target language, and a sentence is built in the target language that expresses the meaning in the translated parse tree. The synthesiser processes the inflected English word form based on the morphological features and traverses the syntactic tree to produce the surface English phrase. An Arabic word can be translated into more than one English word, as shown in Table 1, which leads to the ambiguity problem in the Arabic-to-English MT system.

Table 1 One-to-many MT transfer

Arabic word	English words
معلم	teacher, instructor, schoolteacher
يدرس	teach, study, examine
طالب	student,require,Taleb, requesting

This problem will be solved in the next part of the proposed hybrid approach, which is the statistical part (EM). The EM part estimates the word translation probabilities for selecting the word translation using a parallel corpus.

B. SMT Component

The EM algorithm is an iterative statistical algorithm which is used to find the maximum probability of sentence translations, and it is the most accurate one [16]. The EM algorithm is used as the SMT component system in this study. The EM algorithm contains two steps, i.e., E-step and M-step. The E step uses current observations and parameters to compute the probability of all possible collocations of words. In this paper, the E step is replaced with the rule-based approach, considering that the E-step in MT often fails because the translation provides different alignments for target language distribution that cannot be sampled with simpler change [14]. In the M step, collocation probability is used to re-estimate word translation probabilities in selecting the translation word. Iterating these steps leads to a convergence that estimates the maximum probability of translation. The following Bayes rule is used

$$arg\ max_{e_s} P_s(e_s \mid a_s) = arg\ max_{e_s}\ p_{col}(e_s)p_w(a_s|e_s). \quad (1)$$

For example, the probability of translating an Arabic sentence "المعلم يُدرس الطالب" into an English sentence the probability of each English word e that can be translated into Arabic a is calculated, as shown in Table 2.

TABLE 2 PROBABILITIES THAT ENGLISH WORD ES WILL BE TRANSLATED E_s ARABIC WORD A_S

(a_s/e_s)	$P_{col}(e_s \mid a_s)$
(ال|the)	1
(معلم|teacher)	0.86
(يدرس|teach)	0.57
(طالب|student)	0.61

Based on the given formula, the collocation of the English words that occur in the highest scoring candidate sequence is selected.

$$arg\ max_{es}\ P_s(e_s \mid a_s) = arg\ max_{es}\ p_{col}(the)p_w(\text{ال}|the) \bullet p_{col}(teacher)p_w(\text{معلم}|teacher) \bullet p_{col}(teach)p_w(\text{يدرس}|teach) \bullet p_{col}(the)p_w(\text{ال}|the) \bullet p_{col}(student)p_w(\text{طالب}|student$$

Combinations of RBMT and SMT are highly competitive in MT quality. The output tends to be grammatical, and the main effect of the combination is an increase in lexical selection quality, which strengthens one of the weak points of pure RBMT systems.

C. Hybridisation Algorithm

Owing to the richness and complexity of the Arabic morphology, the word ordering and ambiguity in Arabic are standard characteristics; they represent hurdles in the MT and thus require comprehensive techniques. A new hybridisation strategy is proposed in this study to integrate the rule-based approach with a statistical algorithm (EM). The combination of two effective approaches obtains more accurate results than other existing approaches. Experiments are conducted in Arabic-to-English translation. The hybrid approach relies on the structure output based on the rule-based component and the statistical system. The architecture of the hybrid rule-based approach with EM MT (HMT) system is illustrated in Figure 5.

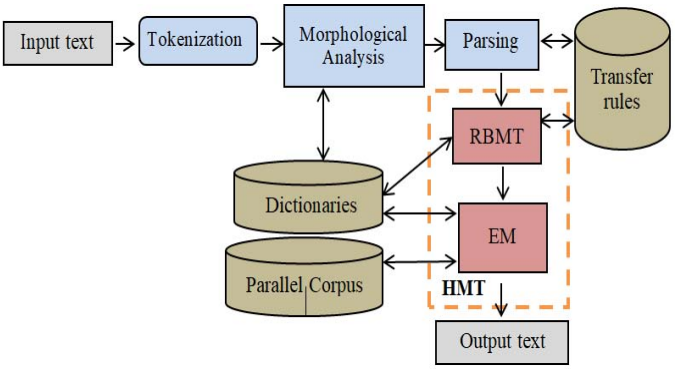


Figure 5 Architecture of the HMT system

During the hybridisation stage, a set of the rules are applied. The statistical model (EM) is then applied to select the most suitable sentence from a number of candidate sentences for reasonable translation. In this study, the best properties of different MT models, which include the linguistic features in SMT, are combined, or the standard RBMT architecture is modified to include statistical knowledge. As mentioned in the previous section, the rule-based approach is combined with the part M of the EM algorithm to address the problems of word ordering and ambiguity. For explanation, the input Arabic sentence "المعلم يُدرس الطالب" is translated using the hybrid approach proposed in this study. The input sentence is tokenized with the first step of the hybrid system by the tokenizer. The morphological analyser analyses and identifies the syntactic information of each word in the source sentence. From the dictionary, the meanings of each word are provided. The sentence is finally split by the parser into a number of sub-sentences, as illustrated in Figure 6. Suitable rule structures for the source sentence and its equivalent target sentence are found in the transfer step to generate an output sentence in the first part of the proposed hybrid system. Figure 6 illustrates the translation processes of the first part of the proposed approach in this paper.

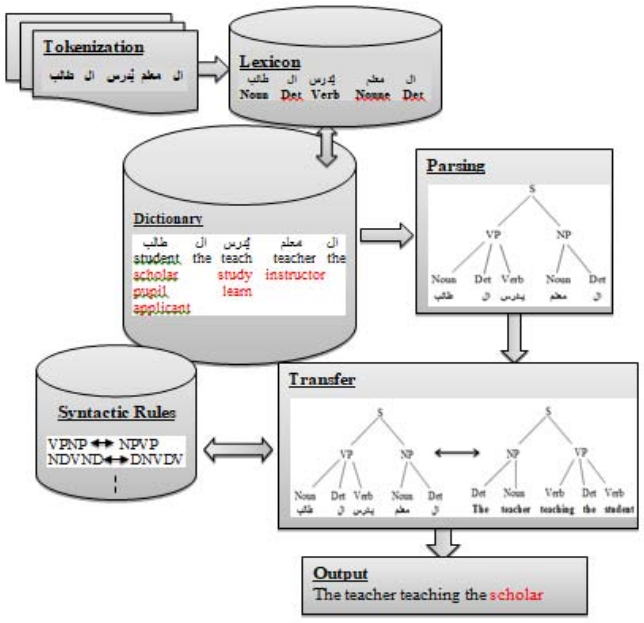


Figure 6 Rule-based part of the hybrid system

The problem of word ordering is solved in the first part of the proposed approach. As shown in Figure 6, the output sentence in the last step is well-ordered because of the availability and robustness of the rule structure of the Arabic sentence and its equivalent English sentence. The algorithm achieves good improvement in the initial stages. However, the translation is not totally correct because the last word in the output sentence is "**scholar**," but for the correct translation, the word should be "**student**." Thus, the output sentence is considered as an ambiguous sentence. This problem will be solved in the next part of the proposed hybrid approach to be discussed in the following paragraphs. Figure 7 shows the second part of the proposed hybrid approach in this study, which represents the EM algorithm.

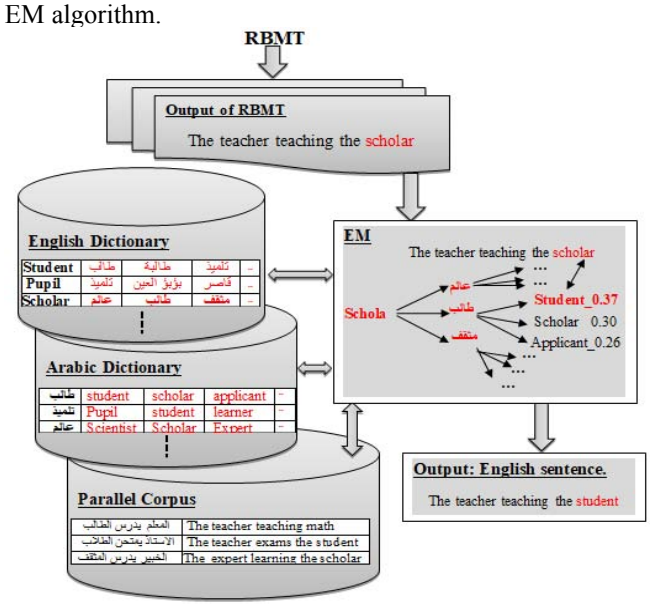


Figure 7 EM algorithm part of the proposed hybrid approach

The English word "scholar" in the output sentence of RBMT may be translated into six Arabic words "**عالم, طالب, علامة, مثقف, طالب يتلقى اعانة, اديب**." In the EM part of the proposed hybrid approach, the frequency of each Arabic word that occurs in the Arabic corpus is collocated using the following formula:

$$P_w(e_i \mid a_i) = \frac{p(e_i)}{total\ p(e_i)} \quad (2)$$

For example, based on Table 3, the probability that the Arabic word "**طالب**" will be translated as the English word "**Student**" can be computed as follows:

$$P_w(e_i \mid a) = P_w(\text{Student} \mid \textbf{طالب}) = 2115/5703 = 0.37$$

The probability that the Arabic word "**طالب**" which will be translated to the English word "Scholar" is:

$$P_w(E_i \mid A) = P_w(\text{SCHOLAR} \mid \textbf{طالب}) = 1711/5703 = 0.30$$

TABLE 3 THE FREQUENCIES OF THE WORD TRANSLATIONS OF THE ARABIC WORD "طالب"

English word	Frequencies
Student	2115
Scholar	1711
Pupil	1220
Applicant	153
Demand	201
Schoolboy	300
Learner	3
	Total 5703

The Arabic words that occur in the highest scoring candidate sequence are then selected. In this case, the Arabic word "**طالب**" is selected. This Arabic word "**طالب**" can also be translated into a number of English words, i.e., **student**, **scholar**, **pupil**, **applicant**, **demand**, **schoolboy**, and **learner.** The frequency of each English word that occurs in English is calculated, and the word with the highest score is selected. The English word "**student**" is selected and replaces the word "scholar" in the output sentence of RBMT. Therefore, the output sentence of the hybrid system is "**The teacher teaches the student**," which is a more accurate translation than that obtained by using only the RBMT system. This system can analyse and identify each word in the source sentence, find a suitable rule structure for the Arabic sentence, and transfer it to its equivalent English sentence. Considering that one Arabic word has many meanings in English, the suitable meaning of a word for the source sentence should be identified. This identification generates candidate sentences with each meaning of a word. Translated English candidate sentences, which are grammatically correct but may be meaningless because of the ambiguity of words, are created. The EM algorithm selects the

most suitable sentence from these candidate sentences. SMT techniques are subsequently utilized. The data-driven RBMT and EM methods are robust. These systems are attractive because they always produce translation regardless of the input string. If an RBMT system does not find a sequence of rules that can be applied successfully to the input, then the EM method performs identification and production. The system finds the probability of the input sentence, which cannot find a suitable rule for it by matching with the corpus directly. The advantages of combining the positive elements of the rule-based approach and the statistical approach to MT are clear, specifically, a combined model has the possibility to be robust, highly accurate, cost effective to build, and adaptable.

IV. EVALUATION RESULTS

In this section, the evaluation results of the developed hybrid approach are described. The proposed approach is tested on the United Nations (Arabic–English) parallel corpus. The same training and test data split is used as in the work of [17], i.e., more than 903 sentence pairs; 632 sentence pairs are used for training and 271sentence pairs are used for testing. Bilingual evaluation understudy (BLEU) is also used to evaluate the proposed system in the evaluation metric score in this paper against two references. Table 4 shows the values of BLEU obtained for the following phrase lengths: 1-*gram*, 2-*gram*, 3-*gram*, and 4-*gram*.

TABLE 4 BLEU EVALUATION RESULTS

Phrase length (n-*gram*)	HMT
1-*gram*	0.89
2-*gram*	0.78
3-*gram*	0.66
4-*gram*	0.59

After performing the analysis toward the RBMT output, comprehensive reordering rules with estimating the maximum probability of translation are observed to play an important role in the translation quality. The HMT system yields good improvement in the BLEU points. Considerable data training makes the HMT output highly accurate. Based on the evaluation results, a good translation can be achieved when RBMT is combined with EM because RBMT solves the word-ordering problem when translating from Arabic to English and the EM method solves the ambiguity problem.

V. CONCLUSION

This study assesses the combination of rule-based and EM approaches. The advantages of transfer rule-based and statistical approaches are utilised to improve the MT from Arabic to English. The employed statistical approaches can improve the MT quality by solving the ambiguity problems that the rule-based approach cannot address, as the ambiguity can only be resolved if we have sufficient, relevant training data to develop the MT systems. However, the rule-based approach is good in handling the word-ordering problem by using a number of effective rules. Evaluation by using the BLEU score indicator shows that the size of the training data affects the statistical model in the EM and HMT systems. Therefore, using a larger training corpus can improve the performance of the HMT system. The hybrid solutions tend to combine the advantages of the individual approaches to achieve an overall better translation. The approach is most useful in addressing one of the greatest rule-based challenges, i.e., translation ambiguity. When a word/phrase can have more than one meaning, statistics can help identify the most suitable option.

REFERENCES

[1] M. R. Costa-jussà, and J.A.R. Fonollosa, 2015. *Latest trends in hybrid machine translation and its applications,* Computer Speech & Language, Volume 32, Issue 1, July 2015, Pages 3-10)

[2] M. R. Costa-jussa, R.E. Banchs, R. Rapp, P. Lambert, K, Eberle and B. Babych. . 2013. *Workshop on Hybrid Approaches to Translation*: Overview and Developments. ACL 2013, 1.

[3] C. Federmann. (2013). *Hybrid machine translation using binary classification models trained on joint, binarised feature vectors*.

[4] V. M. Sánchez-Cartagena, J. A. Pérez-Ortiz and F. S´anchez-Mart´ınez, 2016. *Integrating Rules and Dictionaries from Shallow-Transfer Machine Translation into Phrase-Based Statistical Machine Translation*. Journal of Artificial Intelligence Research 55:17–61.

[5] M. R Costa-Jussà,., C. A. Henríquez, & R. E. Banchs, 2012. *Evaluating indirect strategies for Chinese-Spanish statistical machine translation*. Journal of artificial intelligence research, 45, 761-780.

[6] I.T. Khemakhem, S. Jamoussi and A. B. Hamadou, 2013. "Integrating morpho-syntactic features in English-Arabic statistical machine translation", Proceedings of the Second Workshop on Hybrid Approaches to Translation", Sofia, Bulgaria, pp. 74-81.

[7] A. Alqudsi, N. Omar, N. K. & Shaker, (2014). *Arabic machine translation: a survey*. Artificial Intelligence Review, 1-24

[8] N. Habash, & H. Hassan, (2012). Special issue on Machine Translation for Arabic: Preface. Machine Translation, 26(1-2), 1-2.

[9] R. Zbib, M. Kayser, S. Matsoukas, J. Makhoul and H. Nader, 2012. Methods for integrating rule-based and statistical systems for Arabic to English machine translation. Machine Translation, 26(1), 67-83.

[10] H.W.Xuan,W.Li,and,G.Y.Tang, 2012. *An Advanced Review of Hybrid Machine Translation (HMT)*. Procedia Engineering, 29, 3017-3022.

[11] A. Eisele, C. Federmann, H. Uszkoreit, H. Saint-Amand, M. Kay, M. Jellinghaus, S. Hunsicker, T. Herrmann,. andY. Chen, 2008. *Hybrid machine translation architectures within and beyond the euromatrix project*. In Proceedings of European Association of Machine Translation (Hamburg), pages 27–34.

[12] A. Hatem and N.Omar (2010) *Syntactic reordering for Arabic-English phrase-based machine translation*. In: Database theory and application, bio-science and bio-technology. Springer Lecture Notes in Computer Science, vol 118. Verlag, Berlin, pp 198–206.

[13] A. Hatem, N. Omar & K.Shaker, (2011) Morphological analysis for rule based machine translation. In Semantic Technology and Information Retrieval (STAIR), 2011 International Conference on (pp. 260-263). IEEE.

[14] J. DeNero and K. Dan, 2007. "*Tailoring word alignments to syntactic machine translation*." Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics. 2007.

[15] F. J. Och, Tillmann, C., & Ney, H. (1999). *Improved alignment models for statistical machine translation*. In 1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora.

[16] A. P. Dempster, N. M. Laird & D. B. Rubin (1977). *Maximum likelihood from incomplete data via the EM algorithm*. Journal of the royal statistical society. Series B (methodological), 1-38.

[17] S. Bangalore, H. Patrick, and K. Stephan. "*Statistical machine translation through global lexical selection and sentence reconstruction*." In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pp. 152-159. 2007.

[18] K. Shaalan, and A. H. Hossny, (2012). "*Automatic rule induction in Arabic to English machine translation framework*." Challenges for Arabic Machine Translation 9 (2012): 135.

[19] E. Mohamed, & F. Sadat, (2015). *Hybrid Arabic–French machine translation using syntactic re-ordering and morphological pre-processing*. Computer Speech & Language, 32(1), 135-144.

[20] D. Taji, S. Khalifa, O. Obeid, F. Eryani, & N. Habash, (2018). *An Arabic Morphological Analyzer and Generator with Copious Features*. In Proceedings of the Fifteenth Workshop on Computational Research in Phonetics, Phonology, and Morphology (pp. 140-150).

[21] M. Alkhatib, & K, Shaalan, (2018). *The Key Challenges for Arabic Machine Translation*. In Intelligent Natural Language Processing: Trends and Applications (pp. 139-156). Springer, Cham.