Journal of Engineering Science and Technology
Vol. XX, No. Y (Year) PPP - QQQ

PERSPECTIVES OF ARABIC MACHINE TRANSLATION

HATEM M. ELSHERIF[1], TARIQ RAHIM SOOMRO[2,]*

[1]The Faculty of Engineering & IT British University in Dubai, UAE,
[2]Department of Computer Science, SZABIST, Dubai Campus, UAE
*Corresponding Author: tariq@szabist.ac.ae

Abstract

As the interest on Arabic Language continue to increase worldwide for several factors including political, technological, social and cultural. The significant increase in the Arabic electronic textual information in conjunction with the boom of using social networking and openness to different cultures provided a huge collective knowledge source. Such high demand and urgent need for effective technologies and tools to process and translate information from/to Arabic motivated the researchers in Arabic Machine Translation (AMT) in both the Western and Arab world. This paper aim to explore AMT approaches, challenges and proposed solutions, providing a survey for the research activity conducted on this field and evolutions of the existing related current MT solutions.

Keywords: Machine translation, Arabic machine translation, Natural language processing, Arabic natural language processing.

1. Introduction

MT is the use of computer system to translate automatically text or speech from one natural human language to another [1] MT considered as sub-field of the computational linguistics. Online machine translators use different approaches for MT, approaches commonly classified as Direct, Rule-based, Interlingua, Transfer, Statistical, Example-based, Knowledge-based, and Hybrid Machine Translation. Traditional MT involves processing the grammatical structure and applying linguistic rules in morphology, syntax and semantics to both source and target languages. With the evolving of the Statistical MT approach relying on large bilingual corpus to learn and use probability to nominate the best similar translation, the need for the complex grammatical rules depreciated [2]. Later the

Abbreviations

AMT	Arabic Machine Translation
ANLP	Arabic Natural Language Processing
ANN	Artificial Neural Network
CA	Classical Arabic
DA	Dialectal Arabic
EBMT	Example-based Machine Translation
HMT	Hybrid Machine Translation
KBMT	Knowledge-Based Machine Translation
MSA	Modern Standard Arabic
MT	Machine Translation
RBMT	Rule-Based Machine Translation
SMT	Statistical Machine Translation

Hybrid MT (HMT) approach introduced to combine the strengths of both statistical and rule-based translation methodologies to reduce their individual weakness.

The AMT considered as a major challenge due to many factors which we will discuss later as the grammar and morphology complexity of the Arabic language [3] Researchers consider Arabic language as one of the most difficult machine processing languages [2]. This paper aim to investigate AMT approaches; challenges; and proposed solutions; focusing on the knowledge-based paradigms of machine translation, in order to summarize the research work conducted on this field discussing challenges and research gap. This paper organized as following: First provide a brief background. Second part, we summarize the AMT approaches and related research. The third part review researches related to AMT systems (Arabic/English and English/Arabic) and their evaluation to the solutions. Finally, we summarize the limitations and future work followed by discussion and conclusion.

2. Background

The following sections provides a background of the Arabic Language and its general characteristics followed by overview of Arabic Machine Processing and Translation summarizing research efforts on investigating the challenges in Arabic MT and related work

2.1. Arabic language

Arabic language is one of United Nations (UN) official languages and ranked from the perspective of language speakers as the sixth language in the world[2] It is spoken by not only millions of Arab world inhabitants of the Middle East and Arab League as their native language, but also used by billions of Muslims all over the world as the language of the Holy Quran, it also influencing many other languages. The Arabic language is Semitic language have three main known types for researchers the Classical Arabic (CA) and the new era Modern Standard Arabic (MSA), that is used commonly in the Arabic media differentiating on style and vocabulary than the Classical Arabic. In Arabic, both of them referred to as al-lugha al-fuSHâ "اللغة الفصحى" [4]. In addition to several regional unofficial dialects with major variations and sub-dialects known as Dialectal Arabic (DA) [5].

The basic Arabic alphabet has twenty-eight letters and eight diacritical marks written in horizontal lines from right to left. In addition to eight symbols that can be used as separate letters or used in combination with some of the basic letters with limited combinations, example Hamza "ء" which can be used as a separate letter or used with Alif "ا" in different combinations "أ,إ" and other letters such as "ؤ,ئ". The Arabic letters has different shapes (allographs) according to its position on the word. Example in Table 1. shows letter "م" different shapes as isolated letter or at the begin of word "مـ" and "ـمـ" in the medial (stem part) of word and "ـم" if at the end of word following letter except a set of specific letters such as "ا,ر,د". In addition to ligatures such as "لا" referring to ("ل" +"ا") [6, 7].

Table 1. The Arabic letters different shapes.

Isolated	Begin of word	Medial	End of word
م	مـ	ـمـ	ـم

2.2. Arabic machine processing and translation

Research in Arabic processing started in the 1970s lagging behind other major languages, but research started effectively in the 1980s and grows faster. Driven by different factor the linguistics technology became very important leading to an increasing research on AMT in the west by research institutes and companies [2].

A very comprehensive study arguing some of the ANLP challenges was conducted by Farghaly and Shaalan [8]. In their study, the phenomena of Arabic diglossia and regional dialect ware discussed illustrating the differences between CA and MSA such as grammar structure, lexicon, and morphology. They proposed, through their study the offered solutions for some of the major obstacles facing researchers in the MT field such as Arabic script, Normalization, Syntactic structure. Limited number of books introduced the challenges in Arabic Language MT such as [3, 5] providing an overview of the problems in Arabic natural language processing and current AMT research and solutions, which reflects the gap in this research domain and the need for more research efforts and public Arabic language resources to improve the current AMT to catch-up with other Languages. In addition related good survey conducted by Alqudsi, Omar, and Shaker study the AMT and Machine translation approach [9]. Other surveys on Arabic Language processing were more specific and subject focused such as morphological analysis [10], Named Entity Recognition and Classification [1].

3. Approaches for Arabic machine processing and translation

MT systems can be classified according to the used approach in two main components: knowledge-base, which depends on linguistic (rule-based) approaches and empirical approaches (data-driven) such as the statistical or machine learning approaches hybrid of both [11]. Each approach may have many sub-approaches Fig. 1 show the most commonly used approached.

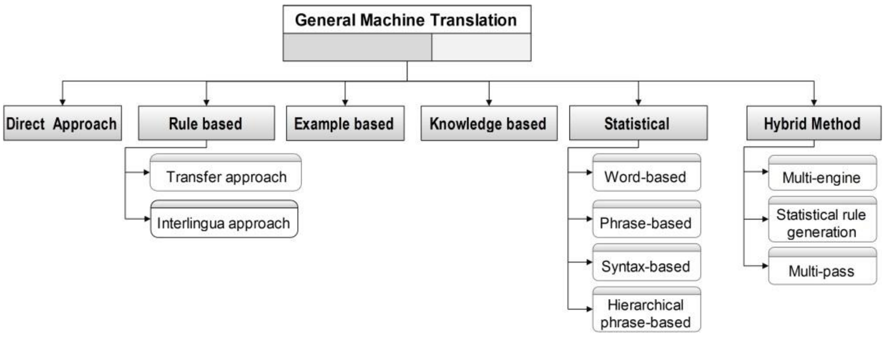

Fig. 1. General Machine Translation Approaches.

3.1. Direct approach

Direct approach is also known as “Dictionary-based machine translation” or “binary translation”, which was mostly used in the first generation of machine translation systems, it depends on the use of dictionaries and designed for a specific pair of languages in one direction translation [12, 13]. Farghaly, stated that the first English to Arabic MT system was developed by Weidner Communication Inc. adapting the direct approach, where the source language was English and the Modern Standard Arabic (MSA) was the target language [14].

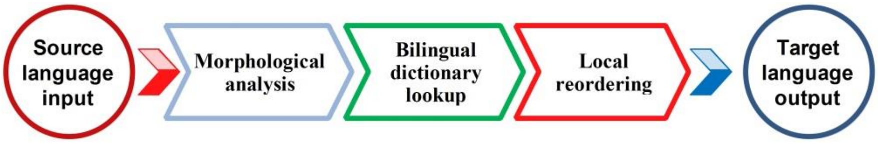

Fig. 2. Direct machine translation approach [9].

The first generation of developed system for machine translation from English to Arabic system adapted the direct approach such as ArabTrans by Apptek in 1990, Al-Mutarjim Al-Araby and Al-Alamiyah other like Al-Nakeel translate from French to both Arabic and English and from English to both French and Arabic [12].

Attia, have studied the Al-Mutarjim Al-Arabey as a direct English to Arabic MT application and found out that the translations weakness because of agreement features different between English and Arabic [15]. Al-Taani & Hailat used Al Mawrid English - Arabic dictionary with the direct MT system approach to translate English to Arabic but the results was 57.3% correct translations [16]. Researchers found out that direct approach lack the capacity of linguistic analysis of the source language and cannot manage the complexity of natural language [14]. Ittycheriah and Roukos proposed in 2007 direct translation model 2 using block style (phrase-pairs) for Arabic-English translation training of millions of parameters based from large corpus on set of blocks improved the direct translation performance [17]. Following studies have offered solutions to reduce this problem by adding linguistic knowledge to the system through applying adequate rules [18], Alawneh & Mohd also offered the use of Hybrid-Based approach adapting combination of both rule-based and example-based approaches [19, 20].

3.2. Rule-based

Rule-Based Machine Translation (RBMT) one of the first techniques used in the AMT, the RBMT use of linguistically knowledge rules and representations which can provide deep analysis for both sentence structure and semantic level but limited to the human written rules [9]. The RBMT has two types of MT systems Interlingua and Transfer approaches, which differentiated according to the level of analysis (as shown in Fig. 3) at the direct approach comes at pyramid bottom with basic analysis level and the Interlingua-based approach comes at the top of the pyramid [21]. Shaalan, presented an overview of the rule-based approach in Arabic clarifying the need for more research efforts in this approech regarding the arabic language to reduce the gap with other mature language technology, And also illustrated its applications in the Arabic language [22].

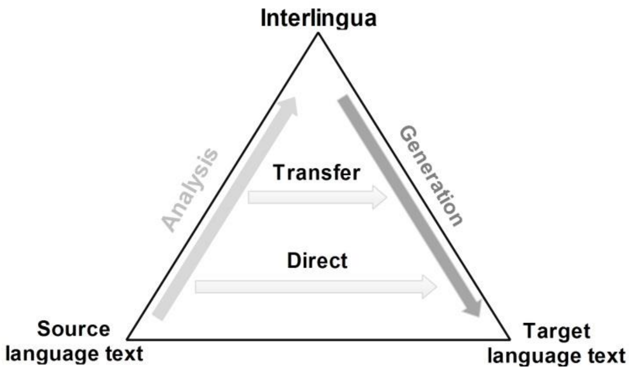

Fig. 3. Bernard Vauquois' pyramid [23].

3.2.1. Interlingua-based approach

Interlingua MT approach is a classic indirect approach work in two stages [24]. As shown in Fig. 4, the first stage analyze the source language sentences into intermediate representation (Interlingua representation), second stage, generates the target sentences text from converting the meaning from the representation. The intermediate representation of the source and target language is language-independent “Universal”, therefor Interlingua MT approach was used in systems that support many languages [25, 26].

The main challenge using the Arabic Interlingua MT approach is designing the Interlingua representation that manage the un-ambiguity while capturing the language semantic structure, which was studied and tested in a limited article that proposed different approaches for different applications. Habash et al., Designed a Conceptual Interlingua for information retrieval [28]. Shaalan et al., used grammar-based generation approach to enhance the Interlingua MT approach by generating grammatically correct Arabic sentence [29]. Shaalan et al., offered a mapping approach to solve the problem of syntactic structure determination reflecting the grammatical structure of the target Arabic sentence [26]. Another approach applied by Adly and Al Ansary in developing and evaluated English to Arabic MT system adapting the Interlingua approach based on the Universal Networking Language [30, 31].

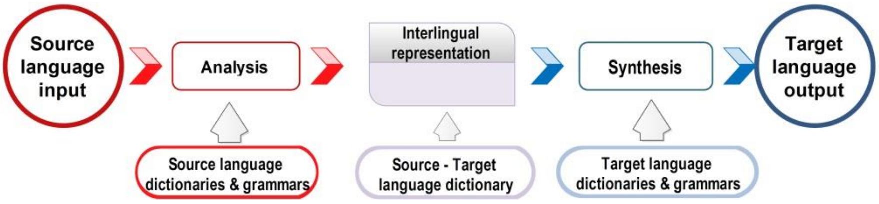

Fig. 4. Interlingua approach (Adapted from [27]).

3.2.2. Transfer-based approach

Transfer-based MT approach follows the same concept of the Interlingua-based approach by using intermediate representation in order to capture the meaning of the source sentence, with the difference from the Interlingua-based approach that the intermediate representation between source and target languages is language-dependent and needs transfer rules.

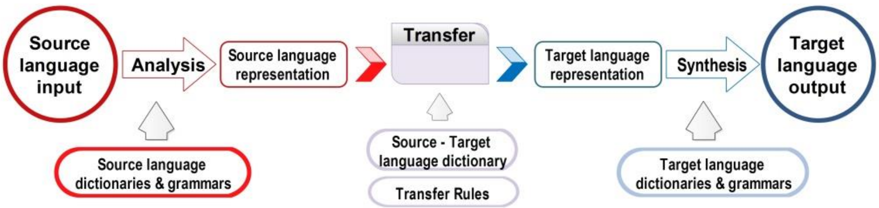

Fig. 5. Transfer-based approach (Adapted from [27]).

The Transfer-based approach has three stages as shown in Fig. 5. Instead of creating one Interlingua representation, it creates two representations for both the source and target language. The first stage analyze the source language sentence to provide syntactically correct abstract representation of the source sentence, second stage involving transforming the source representation to the target language representation, last stage morphological and syntactic generator constrict the target sentence [27, 32]. Researchers applied the Transfer-based approach in MT automating the translations of English Noun Phrases into [33, 34] Arabic .

Shaalan et al., introduced the use of Transfer-based approach developing English-Arabic bi-directional MT system for rapid deveoplment of agricultural expert systems, the system was evaluated by Bi-Lingual Evaluation Understudy (BLEU) using agricultural related set of gold standard parallel English-Arabic [35]. Hatem et al., proposed improvments to transfer-based approach using the morphological analysis [36]. Al Dam & Guessoum, introduced another new approach to improve the English-to-Arabic MT transfer module using Artificial

Neural Networks (ANN) by automatically learn the relation between source and target language structures training the ANN model by using corpus delivering promising evaluation results [37].

3.3. Example-based (EBMT)

Example-based approach or “Memory-based” use MT by analogy without the need to linguistic knowledge relying on the use of large bilingual corpus with parallel texts at real time. The EBMT inspired from the idea of how a human brain process language by analogy, proposing that human sentence translation start by finding sentence and words from the source language and compare it to a memorized similar sentence or word sequences (segments) from the target language before composing these parts reconstructing the target sentence [38, 39]. The EBMT (as shown in Fig. 6) consists of three stages: first identifying the largest phrase of the source sentence that match the example-base “Bi-Lingual Corpus”, which called “Matcher”. Second stage involves finding translation of the matched phrase “Identification Module”. Last stage is constructing the sentence by combining the sentence parts into one similar sentence “Recombination Module” [40].

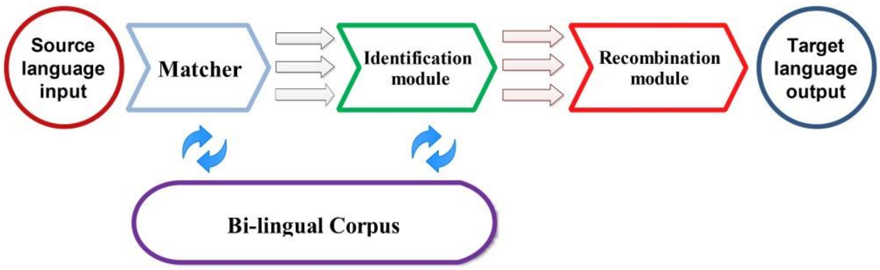

Fig. 6. Example-based approach.

Guidere, refer to corpus-based MT approach as an extension of the Example-based approach using the proposed approach for French-Arabic MT [41]. Cavalli-Sforza & Phillips describe the use of morphology to improve the Example-Based MT translation quality from Arabic to English regardless of the corpora size and they argue that the morphological complexity of the Arabic language makes the number of words occurrence low because of the many surface forms for one word and will by difficult to cover even by using large corprus, using morphology can translate the Arabic language to more generalized class that can be largely correct, when matched against corpora providing better results than the Statistical MT, when little data are available [42]. Bar and Dershowitz proposed the use of Arabic semantic equivalents in the the Example-Based approach translating from Arabic to English using single-word equivalent and multi word paraphrases, founding that using contextual synonyms in EBMT has interesting potentials regardless of the size of the corpus [43, 44]. El-Shishtawy and El-Sammak, discusses the use of template-based technique syntactic matching instead of string-based in the Example-based approach, the proposed technique improve the translation accuracy and reduces the needed corpus size [45].

3.4. Knowledge-based (KBMT)

Knowledge-Based Machine Translation (KBMT) extend the MT capabilities aiming to increase the translation quality by reduce ambiguity as one of the major MT problems by incorporating the real-world knowledge to increases the deep understanding of text meaning. this approach can be used in the process of interpretation of text using knowledge bases as reference providing language-independent representation “meaning” of sentence segment [9, 46]. While researchers claim that its proved that realistic large (KBMT) systems require a huge knowledge about language and about the world [45]. Shaalan and Hossny, propose another approach of using knowledge bases to learn translation rules form positive and negative examples using Inductive Programming Languages (IPL) using a small amount of parallel sentences concluding the advantage of KBMT approach comparing to the corpus-based approaches [3, 47]. While the use of world electronic knowledge base resources such as Wikipedia and WordNet in different research areas in ANLP is increasing such as information retrieval [48, 49], Text Mining [50], Arabic Corpora [51] and Named entity recognition [52, 53].

3.5. Statistical Machine Translation (SMT)

SMT is motivated by information theory and categorized as data-driven approach [54]. it treats translation process as machine learning problem, by examining large sized bilingual corpus using statistical algorithms it automatically learn statistical parameters in order to predict the probability of the most likely (similar) translation [55] as shown in Fig. 7. The SMT do not require linguistic knowledge instead, it learns the rules form the corpus. The SMT approach provide more natural translation and even similar translation to non-trained sentences comparing to the Literal translation delivered by the Rule-Base approaches but can often make obvious errors.

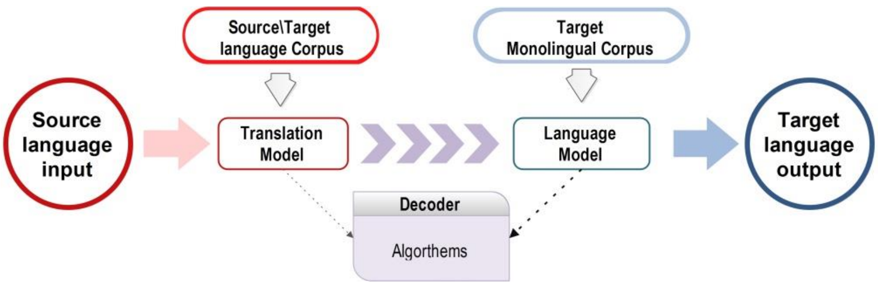

Fig. 7. Statistical machine translation approach (Adapted from [54]).

The general SMT approach consists from three main components, the Translation model that use the parallel corpus to estimate the translation probabilities, the Language model, which use the monolingual corpus to estimate the target probabilities and MT decoder that uses different methods to reduce the noise limiting the search space and adequate quality.

Most research in AMT using the Statistical approach (SMT) focused on the translation from Arabic to English and other languages [11]. Literatures survey for articles in SMT categorized by the used approach and source/target language summary (as shown in Table 2) conclude that majority of contribution of this research filed focused on Phrase-based approach translating from Arabic-English. Very limited research introduced Arabic-Chinese [56, 57].

Table 2. Statistical machine translation researches.

Approach	Arabic- English	English-Arabic
Word-based	[58, 59]	[60]
Phrase-based	[61-66]	[11, 67, 68]
Hierarchical phrase-based	[69, 70]	
Syntax-based	[71]	[72]

3.6. Hybrid method

Hybrid machine translation (HMT) combined the strengths of both statistical and rule-based translation methodologies. While statistical MT lacks the grammatical structure resulting often to ungrammatical sentences, the Rule-based MT lacks the needed lexical coverage. Hybrid architectures intend to provide a better translation getting the best of both paradigms [73]. In general, the HMT approach applied in two ways, first translating by Rule-Based then applying statistics MT to improve the translation, second improving the statistics MT translation by applying post and pre-processing rules.

Costa-jussà & Fonollosa presented in their work a good survey of the latest leading research and progress in HMT approach [74]. Several research on the AMT proved that the use of HMT improve the translation quality, Matusov et al., presented a MT systems that combine five MT system (Multi-engine) to translate from Arabic-to-English with the goal of improving translation quality achieving (55) BLEU score [75]. Mohamed & Sadat presented the HMT approach by introducing the use of morphological rule with SMT to reduces the Arabic morphology level to a closer level to French in translating form Arabic to French providing better results [76]. Habash et al., conduct an evaluation of multiple system for translation form Arabic to English using the HMT approach [77]. Other researches use different combinations approach of statistical models; such as the increasing trend studying the use of Neural Network in HMT approach [78]; Also researchers in Arabic dialect continue to use the HMT approach trying to solve different problem in this area such as [79 - 81]

4. Arabic machine translation systems

4.1. Arabic-to-English MT

Translating form Arabic-to-English researches can be classified in two main groups. The first, focus on proposing or/and evaluating online MT systems such as the work of Farghaly et al., introducing SYSTRAN MT system [82]. Another study by Farghaly comparing three MT systems using different approaches SYSTRAN as transfer MT approach, Google as statistical MT approach and AppTek as hybrid MT approach explaining the improvements in the AppTek translation due to incorporating more futures such as Named entity recognition, Arabic dialects and

Arabic speech recognition [14]. Another MT system Npae-Rbmt, which is adapting the transfer-based approach to translate Arabic noun phrases into English resulting on better general score (94.6%) comparing to Google (80.9%) and Systran (60.2%) [34]. Y Salem et al., introduced UniArab as conceptual in progress work for a developing universal MT system form Arabic-to-English using roles and reference Grammar (Linguistic Model), which is limited to the implemented roles and lack dealing with complex and large sentences [83 - 86].

Other researchers conducted studies evaluating online MT systems. Izwaini, study evaluated three systems (Google, Sakhr, and Systran) evaluating and analyzing it's translation results summarizing the problems and providing recommendations for each system [87]. Kadhim et al., conducted an evaluation translating Arabic news headlines from three online sources into English between Google Translate and Babylon adapting criteria of Hutchins and Somers, using 28 experienced professionals whose native language is Arabic to assess the outputs using questionnaire evaluating the two systems translation quality. The study concludes that both systems has the same clarity (80%) while google translation scored higher accuracy of (77.5%) than Babylon(75%) on the other hand for style criteria Babylon scored (75%) comparing to Google (70%) [88]. The study used a very small sample (40 Arabic news headlines) and only used human evaluation methodology. Another study by Hadla et al, conducted an evaluation translating Arabic to English comparing between Google Translate and Babylon using a constructed two reference professional human translations dataset of 1033 Arabic sentences categorized into four sentence functions (declarative, interrogative, exclamatory, and imperative), adapting Bilingual Evaluation Understudy (BLEU) to evaluate MT quality. The study concluded that in general Google precision values (0.45) better than Babylon (0.4) but Babylon are more effective in translating Arabic exclamation sentences (0.37) than google (0.34) [89]. The study weakness is using relatively small corpus.

Second category of research in MT, proposing solutions for a specific Arabic-to-English translation challenges such as, Arabic word segmentation and tagging [90, 91], Arabic Named Entity recognition and extraction [92 - 96], grammar checker [97], morphological analysis [10, 36, 98] and Dialectal Arabic [99 - 101].

4.2. English-to-Arabic MT

Al-Kabi et al., conducted a comparison between Google Translate and Babylon evaluating translation from English to Arabic adapting BLEU method for evaluation using 100 sentences categorized as following (types, past, present, future, imperative, passive, conditional and questions) and 300 popular English sentences. The study concludes that google is better than Babylon in translating from English to Arabic, the evaluation overall translation precision results for Google scored (0.314), while Babylon scored (0.194) in the corpus set. Regarding the English popular sentences Google scored (0.44) and Babylon scored (0.12) [102].

Akeel and Mishra introduced the using of ANN (Feed-Forward Back-Propagation) adapting the rule-based approach, translating from English into Arabic by learning the meaning of words and linguistic features. The system achieve n-gram blue score(0.6), METEOR score (0.82) and F-measure (0.84) [103]. the proposed solution limited to simple sentences and miss translating

English words that have more than one meaning in Arabic lacking the right morphological analysis capabilities. Researchers introduced UNIARAB MT system translating from English to Arabic evaluating its translation comparing to Google, Tarjim and Babylon adapting three evaluation methods concluding that it performed better on all metrics [30, 31, 104].

5. Discussion and Limitation

AMT comparing to other languages needs massive organized research efforts to achieve the maturity required to catch-up with other Languages. The researcher encountered difficulty to find and classify resources due to the very limited surveys and research communities that collect, classify and maintain the AMT research work except for limited academic institutions research groups leading to a very scattered researches hindering the development of the AMT collective knowledge. In addition to the Black box approach used in the commercial application and the lack of access to Arabic corpus resources and tools for researches due to availability or cost. Which reflects the gap in this research domain and the need for more research efforts and public Arabic language resources to improve the current Arabic MT. Research work evaluating the online MT is limited to the free and open solutions only using different corpus and different evaluation methodologies resulting on different outcomes. AMT is the core of ANLP many research paradigms and application, this study is limited to the perspective of AMT approach and more focused on the Arabic/English and English/Arabic MT systems with limited reference to other languages.

6. Conclusion and Future work

The study concludes that there is an urgent need for a standard complex well tested corpus and evaluation methodology to be commonly recognized by the AMT research community to measure the quality, clarity and performance of the AMT system. While the Rule-Based Machine Translation (RBMT) approaches tend to be the most appropriate approach for Arabic-to-English MT due to the ANLP challenges, SMT specially the Phrase-based approach introduced many solutions leveraged by the increasing usage of online MT system, on the other hand we believe that the future of AMT systems relies on the adaptation of the hybrid approach specially with the increasing usage of Dialectal Arabic in the electronic social media channels. The HMT approach also leveraged by the advantages of the SMT approaches; it is reducing the ambiguity problem in translating from English to Arabic. Future work efforts should survey and categorize the research work from the perspective of AMT applications to guide the research community to the gap in each paradigm.

References

1. Shaalan, K. (2013). A survey of named entity recognition and classification. *Computational Linguistics*, 40(2), 471-510.
2. Zughoul, M.R.; and Abu-Alshaar, A.M. (2005). English/Arabic/English machine translation: A historical perspective. *Meta: Journal des Traducteurs*, 50(3), 1022-1041.

3. Soudi, A.; Farghaly, A.; Neumann, G.; and Zbib, R. (2012). *Challenges for Arabic machine translation*. Amsterdam, Netherlands: John Benjamins Publishing.
4. Ryding, K.C. (2005). *A Reference Grammar of Modern Standard Arabic*, Cambridge, UK: Cambridge University Press.
5. Habash, N.Y. (2010). *Introduction to Arabic Natural Language Processing*, Morgan & Claypool.
6. Rafea, A.A.; and Shaalan, K.F. (1993). Lexical analysis of inflected Arabic words using exhaustive search of an augmented transition network. *Software-Practice and Experience*, 23(JUNE), 567-588.
7. Habash, N.; Soudi, A.; and Buckwalter, T. (2007). On Arabic transliteration, In *Arabic Computational Morphology: Knowledge-Based and Empirical Methods*, Springer, 15-22.
8. Farghaly, A.; and Shaalan, K. (2009). Arabic natural language processing: Challenges and solutions. *ACM Transactions on Asian Language Information Processing*, 8(4), 1-22.
9. Alqudsi, A.; Omar, N.; and Shaker, K. (2014). Arabic machine translation: A survey. *Artificial Intelligence Review*, 42, 549-572.
10. Al-Sughaiyer, I. a.; and Al-Kharashi, I. a. (2004). Arabic morphological analysis techniques: A comprehensive survey. *Journal of the American Society for Information Science and Technology*, 55(3), 189-213.
11. Zbib, R.; and Badr., I. (2012). Preprocessing for English-to-Arabic statistical machine translation, In *Challenges for Arabic Machine Translation*, Amsterdam, Netherlands: John Benjamins Publishing, 95-107.
12. Zantout, R.; and Guessoum, A. (2000). Arabic machine translation: A strategic choice for the Arab World. *Journal of King Saud University - Computer and Information Sciences*, 12, 117-144.
13. Hutchins, J. (2003). Machine translation: General overview, In *The Oxford Handbook of Computational Linguistics*, Oxford University Press, 501-511.
14. Farghaly, A. (2010). Arabic machine translation: A developmental perspective. *International Journal on Information and Communication Technologies*, 3(3), 3-10.
15. Attia, M.A.-A. (2002). *Implications of the agreement features in machine translation*. Ph.D. Dissertation, Al-Azhar University, Egypt.
16. Al-Taani, A.T.; and Hailat, E.M. (2005). A direct English-Arabic machine translation system. *Information Technology Journal*, 4(3), 256-261.
17. Ittycheriah, A.; and Roukos, S. (2007). Direct translation model 2, In *Computational Linguistics*, 57-64.
18. Abu Shquier, M.M.; and Sembok, T.M.T. (2008). Word agreement and ordering in English-Arabic machine translation, In *Proceedings - International Symposium on Information Technology 2008, ITSim*, 1-8.
19. Abu Shquier, M.M.; Al Nabhan, M.M.; and Sembok, T.M. (2010). Adopting new rules in rule-based machine translation, In *Computer Modelling and Simulation (UKSim), 2010 12th International Conference on IEEE*, 62-67.
20. Alawneh, M.; and Mohd, T. (2011). Handling agreement and words reordering in machine translation from English to Arabic using hybrid-based

systems. *Journal of Computer Science*, 11(6), 93-97.

21. Dorr, B.J.; Jordan, P.W.; and Benoit, J.W. (1999). A survey of current paradigms in machine translation. *Advances in Computers*, 49, 1-68.
22. Shaalan, K. (2010). Rule-based approach in Arabic natural language processing. *International Journal on Information and Communication Technologies*, 3(3), 11-19.
23. Vauquois, B. (1968). A survey of formal grammars and algorithms for recognition and transformation in mechanical translation, In *IFIP Congress (2)*, 1114-1122.
24. Hutchins, J. (1986). Strategies and methods since the mid 1960s, In *Machine Translation: past, present, future*, Chichester: Ellis Horwood.
25. Sloculn, J. (1985). A survey of machine translation: Its history, current status, and future prospects. *Computational Linguistics*, 11(1), 1-17.
26. Shaalan, K.; Abdel, A.; Rafea, A.; and Baraka, H. (2006). Mapping interlingua representations to feature structures of Arabic sentences, In *The Challenge of Arabic for NLP/MT*, London, 149-159.
27. Hutchins, J. (1986). Problems, methods, and strategies, In *Machine Translation: past, present, future*, Chichester: Ellis Horwood.
28. Habash, N.; Mah, C.; Imran, S.; Calistri-yeh, R.; and Sheridan, P. (1998). Design, construction and validation of an Arabic-English conceptual interlingua for cross-lingual information retrieval. *Information Retrieval*, 107-112.
29. Shaalan, K.; Monem, A.A.; and Rafea, A. (2003). Generating Arabic text from interlingua, In *Proceedings of the 2nd workshop on computational approaches to Arabic script-based languages (CAASL-2)*, Stanford, CA: Linguistic Institute, 137-144.
30. Adly, N.; and Al Ansary, S. (2010). Evaluation of Arabic machine translation system based on the universal networking language. *Natural Language Processing and Information Systems*, 5723, 243-257.
31. Alansary, S. (2011). Interlingua-based machine translation systems: UNL versus other interlinguas, In *11th International Conference on Language Engineering, Ain Shams University*, Cairo, Egypt.
32. Abdel Monem, A.; Shaalan, K.; Rafea, A.; and Baraka, H. (2008). Generating Arabic text in multilingual speech-to-speech machine translation framework. *Machine Translation*, 22(4), 205-258.
33. El-aini, S.K.; Shaalan, K.; Rafea, A.; Moneim, A.A.; and Baraka, H. (2004). Machine translation of English noun phrases into Arabic. *International Journal of Computer Processing of Languages*, 17(02), 121-134.
34. Shirko, O.; Omar, N.; Arshad, H.; and Albared, M. (2010). Machine translation of noun phrases from Arabic to English using transfer-based approach. *Journal of Computer Science*, 6(3), 350-356.
35. Shaalan, K.; Hendam, A.; and Rafea, A. (2012). Rapid development and deployment of bi-directional expert systems using machine translation technology. *Expert Systems with Applications*, 39(1), 1375-1380.
36. Hatem, A.; Omar, N.; and Shaker, K. (2011). Morphological analysis for rule based machine translation, In *2011 International Conference on Semantic*

Technology and Information Retrieval, Putrajaya, Malaysia: IEEE, 260-263.

37. Al Dam, R.; and Guessoum, A. (2010). Building a neural network-based English-to-Arabic transfer module from an unrestricted domain, In *Machine and Web Intelligence (ICMWI), 2010 International Conference on. IEEE*, 94-101.
38. Nagao, M. (1984). A framework for a mechanical translation between Japanese and English by analogy principle, In *Artificial and Human Intelligence*, NATO Publications., 351-354.
39. Hutchins, J. (2005). Example-based machine translation: A review and commentary. *Machine Translation*, 19(3-4), 197-211.
40. Zitouni, I. (2014). *Natural language processing of semitic languages*, Springer.
41. Guidere, M. (2002). Toward corpus-based machine translation for standard Arabic. *Translation Journal*, 6(1).
42. Cavalli-Sforza, V.; and Phillips, A.B. (2012). Using morphology to improve example-based machine translation: The case of Arabic-to-English translation, In *Challenges for Arabic Machine Translation*, Amsterdam, Netherlands: John Benjamins Publishing, 23-48.
43. Bar, K.; and Dershowitz, N. (2011). Using verb paraphrases for Arabic-to-English example-based translation, In *Machine Translation and Morphologically-rich Languages: Research Workshop of the Israel Science Foundation, University of Haifa, Israel*.
44. Bar, K.; and Dershowitz, N. (2012). Using semantic equivalents for Arabic-to-English example-based translation, In *Challenges for Arabic Machine Translation*, Amsterdam, Netherlands: John Benjamins Publishing, 49-71.
45. El-Shishtawy, T.; and El-Sammak, A. (2014). The best templates match technique for example based machine translation. *arXiv preprint arXiv:1406.1241*, (1), 1-11.
46. Hutchins, W.J.; and Somers, H.L. (1992). Some other systems and directions of research, In *An introduction to machine translation*, London: Academic Press, 313-333.
47. Shaalan, K.; and Hossny, A.H. (2012). Automatic rule induction in Arabic to English machine translation framework, In *Challenges for Arabic Machine Translation*, Amsterdam, Netherlands: John Benjamins Publishing, 135-154.
48. Mohamed I. Eldesouki.; Arafa, W.; Darwish, K.; and Gheith, M. (2011). Using wikipedia for retrieving arabic documents, In *Proceedings of Arabic Language Technology International Conference (ALTIC)*, Alexandria, Egypt.
49. Saad, M.; Langlois, D.; and Smaïli, K. (2013). Extracting comparable articles from Wikipedia and measuring their comparabilities. *Procedia - Social and Behavioral Sciences*, 95, 40-47.
50. Saad, M.; and Ashour, W. (2010). Arabic morphological tools for text mining, In *6th International Conference on Electrical and Computer Systems (EECS'10), Nov 25-26, 2010, Lefke, Cyprus*, 25-26.
51. Zaghouani, W. (2014). Critical survey of the freely available Arabic corpora current situation of the freely available, In *Proceedings of the Workshop on Free/Open-Source Arabic Corpora and Corpora Processing Tools Workshop Programme,LREC*.
52. Abouenour, L.; Bouzoubaa, K.; and Rosso, P. (2010). Using the Yago

ontology as a resource for the enrichment of named entities in Arabic WordNet, In *Proceedings of The Seventh International Conference on Language Resources and Evaluation (LREC 2010) Workshop on Language Resources and Human Language Technology for Semitic Languages*, Valletta, Malta, 27-31.

53. Attia, M.; Toral, A.; Tounsi, L.; Monachini, M.; and Genabith, J. Van. (2005). An automatically built named entity lexicon for Arabic, In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC 2010)*, Valletta, Malta, 3614-3621.
54. Hassan, H.; and Darwish, K. (2014). Statistical machine translation, In *Natural Language Processing of Semitic Languages*, Springer, 199-220.
55. Lopez, A. (2008). Statistical machine translation. *ACM Computing Surveys*, 40(3), 1-49.
56. Habash, N. (2009). Improving Arabic-Chinese statistical machine translation using English as pivot language. *Computational Linguistics*, (March), 173-181.
57. Ghurab, M.; Zhuang, Y.; Wu, J.; and Abdullah, M.Y. (2010). Arabic-Chinese and Chinese-Arabic phrase-based statistical machine translation systems. *Information Technology Journal*, 9(4), 666-672.
58. Sadat, F.; and Habash, N. (2006). Combination of Arabic preprocessing schemes for statistical machine translation, In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL*, New York, USA: Association for Computational Linguistics, 49-52.
59. Mansour, S. (2010). MorphTagger : HMM-based Arabic segmentation for statistical machine translation, In *Proceedings of the 7th Internationa Workshop on Spoken Language Translation*, Paris, France, 321-327.
60. El Kholy, A.; and Habash, N. (2012). Orthographic and morphological processing for English-Arabic statistical machine translation. *Machine Translation*, 26(1-2), 25-45.
61. Koehn, P.; Axelrod, A.; Mayne, A.B.; Callison-Burch, C.; Osborne, M.; and Talbot, D. (2005). Edinburgh system description for the 2005 IWSLT speech translation evaluation, In *Proceedings of the International Workshop on Spoken Language Translation*, 68-75.
62. Al-onaizan, Y.; and Papineni, K. (2006). Distortion models for statistical machine translation, In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, Sydney: Association for Computational Linguistics, 529-536.
63. Diab, M. (2009). Second generation AMIRA tools for Arabic processing: Fast and robust tokenization , POS tagging , and base phrase chunking, In *Proceedings of the Second International Conference on Arabic Language Resources and Tools*, 285-288.
64. Green, S.; Galley, M.; and Manning, C.D. (2010). Improved models of distortion cost for statistical machine translation, In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, 867-875.
65. Mansour, S.; and Ney, H. (2012). Arabic-segmentation combination

strategies for statistical machine translation. *LREC-2012*, 3915-3920.

66. Alkhouli, T.; Guta, A.; and Ney, H. (2014). Vector space models for phrase-based machine translation, In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, Doha, Qatar: Association for Computational Linguistics, 1-10.
67. Elming, J.; and Habash, N. (2009). Syntactic reordering for English-Arabic phrase-based machine translation, In *Proceedings of the EACL 2009 Workshop on Computational Approaches to Semitic Languages - Semitic '09*, Athens, Greece: Association for Computational Linguistics, 69-77.
68. Al-Haj, H.; and Lavie, A. (2012). The impact of Arabic morphological segmentation on broad-coverage English-to-Arabic statistical machine translation. *Machine Translation*, 26(1-2), 3-24.
69. Marton, Y.; Chiang, D.; and Resnik, P. (2012). Soft syntactic constraints for Arabic-English hierarchical phrase-based translation. *Machine Translation*, 26(1-2), 137-157.
70. Huck, M.; Vilar, D.; Stein, D.; and Ney, H. (2011). Advancements in Arabic-to-English hierarchical machine translation, In *Preoceeding of the 15th Conference of the European Association for Machine Translation*, Leuven, Belgium: European Association for Machine Translation, 273-280.
71. Khalilov, M.; and Fonollosa, J.A.R. (2011). Syntax-based reordering for statistical machine translation. *Computer Speech and Language*, 25(4), 761-788.
72. Galley, M.; Graehl, J.; Knight, K.; Marcu, D.; DeNeefe, S.; Wang, W.; and Thayer, I. (2006). Scalable inference and training of context-rich syntactic translation models, In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, 961-968.
73. Hunsicker, S.; Yu, C.; and Federmann, C. (2012). Machine learning for hybrid machine translation, In *Proceedings of the Seventh Workshop on Statistical Machine Translation*, Association for Computational Linguistics, 312-316.
74. Costa-jussà, M.R.; and Fonollosa, J. a. R.R. (2014). Latest trends in hybrid machine translation and its applications. *Computer Speech & Language*, 32(1), 3-10.
75. Matusov, E.; Leusch, G.; and Ney, H. (2009). Learning to combine machine translation systems, In *Learning Machine Translation*, Cambridge, UK: MIT Press, 257-276.
76. Mohamed, E.; and Sadat, F. (2015). Hybrid Arabic-French machine translation using syntactic re-ordering and morphological pre-processing. *Computer Speech & Language*, 32(1), 135-144.
77. Habash, N.; Dorr, B.; and Monz, C. (2009). Symbolic-to-statistical hybridization: Extending generation-heavy machine translation. *Machine Translation*, 23(1), 23-63.
78. Zantout, R.; and Guessoum, A. (2015). Obstacles facing Arabic machine translation: building a neural network-based transfer module, In *Papers in Translation Studies*, Cambridge, UK: Cambridge Scholars Publishing, 229-253.
79. Bakr, H.M.A.; Shaalan, K.; and Ziedan, I. (2008). A hybrid approach for

converting written Egyptian colloquial dialect into diacritized Arabic, In *The 6th International Conference on Informatics and Systems, (INFOS2008)*, Cairo, Egypt: Cairo University, 27-33.

80. Sawaf, H. (2010). Arabic dialect handling in hybrid machine translation, In *Proceedings of the 9th Conference of the Association for Machine Translation in the Americas (AMTA)*, Denver, Colorado.
81. Tachicart, R.; and Bouzoubaa, K. (2014). A hybrid approach to translate Moroccan Arabic dialect, In *Intelligent systems: Theories and applications (sita-14), 2014 9th international conference*, Rabat, Morocco: IEEE., 7-8.
82. Farghaly, A.; Senellart, J.; Systran, S. a.; and Cimetiere, R. (2003). Intuitive coding of the Arabic lexicon, In *SYSTRAN, MT, Summit IX Workshop, Machine Translation for Semitic Languages: Issues and Approaches*.
83. Salem, Y.; Hensman, A.; and Nolan, B. (2008). Towards Arabic to English machine translation. *ITB Journal*, (17), 20-31.
84. Salem, Y.; Hensman, A.; and Nolan, B. (2008). Implementing Arabic-to-English machine translation using the role and reference grammar linguistic model, In *Proceedings of the Eighth Annual International Conference on Information Technology and Telecommunication (IT&T 2008)*, Galway, Ireland.
85. 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, Institute of Technology Blanchardstown, Ireland.
86. Nolan, B.; and Salem, Y. (2009). UniArab: An RRG Arabic-to-English machine translation software, In *Proceedings of The 2009 International Conference on Role and Reference Grammar*, University of California, Berkeley, USA, 243-270.
87. Izwaini, S. (2006). Problems of Arabic machine translation: Evaluation of three systems, In *Proceedings of the International Conference on the Challenge of Arabic for NLP/MT*, London: The British Computer Society (BSC), 118-148.
88. Kadhim, K. a.; Habeeb, L.S.; Sapar, A.A.; Hussin, Z.; and Abdullah, M.M.R.T.L. (2013). An evaluation of online machine translation of Arabic into English news headlines: Implications on students' learning purposes. *Turkish Online Journal of Educational Technology*, 12(2), 39-50.
89. Hadla, L.S.; Hailat, T.M.; and Al-kabi, M.N. (2014). Evaluating Arabic to English machine translation. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 5(11), 68-73.
90. Isbihani, A. El. (2006). Morpho-syntactic Arabic preprocessing for Arabic-to-English statistical machine translation. *Computational Linguistics*, 15-22.
91. Gahbiche-Braham, S.; Bonneau-Maynard, H.; Lavergne, T.; and Yvon, F. (2012). Joint segmentation and POS tagging for Arabic using a CRF-based classifier, In *LREC*, 2107-2113.
92. Elsebai, A.; Meziane, F.; and Belkredim, F. (2009). A rule based persons names Arabic extraction system. *Communications of the IBIMA*, 11(August), 53-59.
93. Shaalan, K.; and Raza, H. (2009). NERA: Named entity recognition for Arabic. *Journal of the American Society for Information Science and Technology*, 60(8), 1652-1663.

94. Zaghouani, W. (2012). RENAR: A rule-based Arabic named entity recognition system. *ACM Transactions on Asian Language Information*, 11(1), 1-13.
95. Zayed, O.H.; and El-Beltagy, S.R. (2012). Person name extraction from modern standard Arabic or colloquial text, In *Informatics and Systems (INFOS), 2012 8th International Conference on IEEE*, 44-48.
96. Zayed, O.; El-Beltagy, S.; and Haggag, O. (2013). A novel approach for detecting Arabic persons' names using limited resources. *Complementary proceedings of 14th International Conference on Intelligent Text Processing and Computational Linguistics, CICLing*, 70, 81-93.
97. Shaalan, K.F. (2005). Arabic GramCheck: A grammar checker for Arabic. *Software-Practice and Experience*, 35(7), 643-665.
98. Sawalha, M.; Atwell, E.; and Abushariah, M.A.M. (2013). SALMA: Standard Arabic language morphological analysis, In *Communications, Signal Processing, and their Applications (ICCSPA), 2013 1st International Conference on IEEE*, 1-6.
99. Habash, N.; Roth, R.; Rambow, O.; Eskander, R.; and Tomeh, N. (2013). Morphological analysis and disambiguation for dialectal Arabic, In *Proceedings of NAACL-HLT*, Atlanta, Georgia: Association for Computational Linguistics, 426-432.
100. Salloum, W.; and Habash, N. (2013). Dialectal Arabic to English machine translation: Pivoting through modern standard Arabic, In *Proceedings of NAACL-HLT*, Atlanta, Georgia: Association for Computational Linguistics, 348-358.
101. Salloum, W.; and Habash, N. (2014). ADAM: Analyzer for dialectal Arabic morphology. *Journal of King Saud University - Computer and Information Sciences*, 26(4), 372-378.
102. Al-Kabi, M.N.; Hailat, T.M.; Al-Shawakfa, E.M.; and Alsmadi, I.M. (2013). Evaluating English to Arabic machine translators Using BLEU. *International Journal of Advanced Computer Science and Applications*, 4(1), 66-73.
103. Akeel, M.; and Mishra, R. (2014). ANN and rule based method for English to Arabic machine translation. *The International Arab Journal of Information Technology*, 11(4), 396-406.
104. Alansary, S.; Nagi, M.; and Adly, N. (2009). The universal networking language in action in English-Arabic machine translation, In *Proceedings of 9th Egyptian Society of Language Engineering Conference on Language Engineering, (ESOLEC 2009)*, 23-24.