07

MACHINE-TRANSLATION APPROACHES AND ITS CHALLENGES IN THE 21ST CENTURY

Laith Salman Hassan Hadla, *Department of English Language, Literature and Translation, Zarqa University, Zarqa, Jordan*

Abstract:

Language is a communication tool for different cultures around the world. Hence, accessibility to other language web documents has always been a concern for information professionals. Consequently, the MT approach, which is a software used to translate a text from one language into another, is widely utilized. It is possible to refer to automatic or immediate translator as compared with human translator it is much quicker. In addition, with the fast change in information and communication technology (ICT), the world has become a small village where individuals can interact with each other via MT. Then MT performs an significant position in personal and business areas and is highly relevant. The translation system can be bilingual or multilingual. Most machine translations are bidirectional, either from origin to target or from destination to source language. If not, it is unidirectional to convert only from the source language into the target language. There are many ways to automatically translate the language. This article focuses on MT's various approaches and also presents several challenges faced in the language translation since English language is evolutionary. This study will also lead to MT by assisting researchers to summarize the significance of the MT system and the MT groups of studies as well as to let lights on the importance of the translation mechanism.

Keywords: *Corpus-based machine translation, Hybrid-based translation, Information and communication technology (ICT), Machine translation (MT), Natural Language Processing (NLP), Rule-based machine translation.*

1. Introduction

In recent times, the technology of information and communication (ICT) has risen quickly. This has influenced the growth of business (Luo & Bu, 2016) and the telecommunications technologies, in which people from many other cultures are able to interact. Languages are thus an important instrument for communication based on their significance in terms of trade, study or relationship with others from other groups and religions (Newmark, 1988). Languages are necessary to deliver emails and comprehend the texts they also receive. It is therefore essential to have a technique of translation that means the exchange of two texts in two distinct codes.

The translation has been needed since the far away past (Newmark, 1988) when human started having journeys for trading, and then for studying. The various deals with various people from diverse cultures that have different languages made it necessary for the government, at that time, to send some chosen individuals to learn those other people's language and culture. This way of interaction was how the human translation started. The human translator must have a huge experience, in that language before starting translating it. The translation itself needs substantial modelling of knowledge about the world, the translation sciences and the languages we translate between them (Macketanz, Avramidis, Burchardt, Helcl, Srivastava, 2017). The less the experience, the worse the translation (Dickins, Hervey, Higgins, 2002), the most important aspects of the human translation are that they are precise and always bilingual.

However, the human translator does not exist every time we need them, and it is not always for free. Also, the privacy of the conversations or the translated messages depends on the translator's ethics and morals.

Nowadays, Translation is a significant activity in people' daily lives (Hatim & Munday, 2004) and it is as old as the idea of the computer (Martin & Jurafsky, 2000) as a result of the Telecommunication Technology which makes the communication much easier, anytime, anywhere and many times for free. Thus, the Computer Science engineers started thinking about making the machine able to translate messages which first introduced to the machine as texts. Unfortunately, translation is a difficult task (Moussallem, Ngomo, Wauer, 2016) because of its dependency on the Natural Language Processing (NLP) and the structure of the languages (Martin & Jurafsky, 2000). The only way to make sense was by using the artificial intelligence because it merges the different required types of sciences during the translation process. A new field of the Machine Learning was to develop algorithms for the Natural Language Processing (NLP) (Indurkhya & Damerau., 2010). These algorithms can be used to analyze the source texts (Indurkhya & Damerau., 2010) into its Parts-Of-Speech (POS) tags to facilitate the translation mechanism. These algorithms are Data Mining (DM) algorithms, such as the POS tools (Ba-Alwi, Albared, & Al-Moslmi, 2017) which are the encoder or decoder in the Neural Machine Translation (NMT) (Almahairi, Cho, Habash, & Courville, 2016). Also, merged with the Data Mining (DM) algorithms in order to help during translation, such as the parsers or Morphological Analyzers tools those used in the Statistical Machine Translation (SMT) (Moussallem & Choren, 2015, Koehn, Och, & Marcu, 2003).

Many types of research are proposing Translations with the help of the machine. Machine Translation is a contemporary research field which means a translation that is providing a fully automated translation. It is distinct from computer-assisted translation, machine-assisted human translation (MAHT) and human-aided machine translation (HAMT) (Hutchins, 1994). Computer aided Translation (CAT) is a language translation method used by a human trader to assist and promote the translation process with computer software (Bowker, 2002). With the help of the machine, MAHT is a human translation. In this machine, there's a program that can check spelling, grammar, or the style of the translation (Hutchins & Somers, 1992). HAMT is a machine that is responsible for translation process with an assistance of the human whenever needed. This human either can be involved while the process "interactive" or can be outside the process "pre-edit" or "post-edit" (Hutchins & Somers, 1992).

Most study discuss about the development of Machine Translation, which covers three types: Statistical Machine Translation (SMT), Machine Translation using Semantic Web, and Neural Machine Translation (NMT). To the best of our knowledge, none of these machines give precise translation like the human translation (Taylor, 2009). Except for the translation of the texts, this recorded in the knowledge base of the machine as assigned phrases which are the full phrases from the source language and its meaning in the target language (Koehn et al., 2003, Brown, Cocke, Pietra, Pietra, Jelinek, & Lafferty, 1990). That is because the words translation has been saved as it is to the knowledge base as translation experts translate it. However, we cannot insert every single phrase and ask the user to translate it. The wheel of the sciences did not stop. The Machine Translations using Semantic Web brings us a new chance for translation using the semantic meanings of the texts (Seo, Il-Sun, Su-Kyung, & Ho-Jin., 2009). However, it is slower than other machines types (Moussallem & Ngomo, 2016) and if the meanings of the homograph words are not added to the ontology file, the translation could be bad.

Some other tools are developed according to much research to help in the Machine Translations area. Some examples of these tools are the Syntax Analyzers such as the parsers (Manning, Surdeanu, Bauer, Finkel, Bethard, & McClosky, 2014, Othman, Shaalan, & Rafea, 2003) and the Morphological analyzers (Zaghal & Sbeih, 2012, Sawalha, Atwell, & Abushariah, 2013, Attia, 2006). These tools are especially used to analyze the text to tell the user, either the human or the Machine, the POS tags (Ba-Alwi et al., 2017) and define the stems of the words. Some other machines build their Syntax analyzers such as the encoder of the NMT. This research is going to provide a summary of the translation itself. Then the research will explore the different types of the Machine Translation also presents several challenges faced in the language translation since English language is evolutionary.

2. Translation

Translation is a very rapid growing activity these days (Hatim & Munday, 2004). It is an active means to transfer the culture and language due to the contact with one other (Newmark, 1988). Its main aspect understands the meaning assigned to the words in the vocabularies of the language (Jakobson, 1959). Any translated text is judged and accepted by the reviewers when it is read as it was meant in its source language, in other words, it is not literally translated, this is because the author may express his or her thoughts and feelings in writing or speaking (Venuti, 2008).

According to the World Bank group (Group, 2004), Modern Standard Classical Arabic is used while writing books, contracts, real articles, such as poets (Group, 2004), and also in the holy Quran. The vernaculars are so hard to be translated because of their dependency on the huge experience in the Arabic Language Sciences and their interference with some other vernaculars and foreign languages. Literal translation, which means translation word by word, is not the best choice, especially when the words meanings alone cannot give us the needed meaning from the sentence at all (Dickins et al., 2002). Newmark stated that (Newmark, 1988) a text may need to be contemplated before being translated in ten different directions, which can be summarized as follow: The idiolect of the author of the Source Language (SL). For this form of document frequent linguistic and lexical use relies on the background and the material objects and specifically references to SL or fifth language societies (i.e., not SL or the target language). Tradition is at that moment impacted by the normal buildings of a document in novels or journals. With regard to their assessed information on the subject and the dialect style, the expectations for the supposed readership communicated to the greatest common factor, since the readership should not be translated (or upgraded). The same applies to TL with the last three marks. Where feasible, whatever is defined or announced, discovered and checked (the benchmark reality), regardless of the SL document and readership aspirations. The translator's private, subjective or social and cultural views and biases include' community fidelity factors' for a translator, which can represent national, political, racial, religious, social, gender, etc. expectations of the interpreter (Newmark, 1988). Dickins et al. (2002), has divided the texts to be translated into five matrices: First is Genre: this is related to the literary, philosophy, thoughts and religion. The second one is Cultural: this is related to idioms, cultural transplantations stories, and proverbs. The third matrix is Semantic matrix: this is related to meaning of synonyms, collocations, and original metaphors. After that Formal: this is related to levels of graphics, sentences, grammar, and Quranic allusions. Last but not least is Varietal: which is related to dialects, sociolects, tones and social registers (Dickins et al., 2002).

On the other hand, the Target Text (TT) that is the translation is not entirely similar to the Source Text (ST) as to its structure or content because of the limitations imposed by the semantic and the formal differences between the Source Language (SL) and the Target Language (TL). The users of the TT evaluate it from three different perspectives: which are functional, structurally and semantically. The functional perspective deals with the text as if it is really done by the author in TT. The structure perspective evaluates the sequence and the arrangement of the ideas in the TT according to the Source Text (ST). The semantic perspective is related to testing the meaning of the TT in comparison with the ST (Kuzenko, 2008).

Newmark (1988) indicated that learning is split into an emotional, informative, vocational, artistical, phatic, metalingual feature according to the required interpretation feature. The expressive function relies on the author's mind and reflects his emotions in many instances. Serious creative literature, lyric poems, brief tales, films or theatrical performances, authoritative accounts, autobios, essay and private letters, and ultimately, memoir and documentary material by ministers or representatives of parties, laws and legal records, science journals, philosophical and academic books published by the respective officials. The informative feature involves described concepts or concepts, often in conventional textbooks. For example, a textbook, a technical report or a newspaper or newspaper post or science paper, a thesis and minutes or a conference schedule are examples of informative writings. The previous four items define it on a scale of different languages: the non-emotive mode of scholarly articles for official purposes, a friendly or casual style with specified techniques for comics, a casual, hot style for popular science or art books and the common, racial, non-technical mode of common journalism.

The public, the recipient is the vocational feature of the language. The connection between the writer and the public lies with all vocative documents, which must be formulated in a language that is immediately publicly understandable. These three kinds are common but usually or binarily discovered. The creative feature is a language that pleases the ears, their real or envisioned voices and their metaphors. The metaphor is a blend of the creative and emotional functions. The phatic function is not used for publishing overseas data but to keep pleasant communication with the receiver. The metalingual function shows the capacity to describe, mark and criticize its characteristics in languages (Newmark, 1988).

Taylor (2009) stated that, many of the challenges while the translation was the great diversity of structures and the process of interpreting the sentence that used a specific structure and then choosing the best structure of the sentence in the target language. The translation was a commercial activity in 1990 (Boucau, 2010), and one of the first Statistical Machine Translations was developed by IBM (Brown et al., 1990) which was intended to automate the core translation task Taylor (2009). There are many types of the translations that are using the technology to aid translation, which will be covered in the following section.

3. Machine-Translation (MT)

The Machine Translation is one of the significant and difficult areas in our everyday lives. Bar-Hilal said MT has become an enterprise of several million dollars (Bar-Hillel, 1960). It is because it needs very complex processes to achieve reasonable results. There are many types of research about building or developing Machine Translation for the English language and many other natural languages. In contrast, similar types of research for the Arabic language are very few and somehow nonexistent. For instances, the previous research papers for the Arabic Language (with other languages like Hebrew) was a Machine-Aided Human Translation (MAHT) or just a Natural Language Processing (NLP) pieces of research to develop a parser which has many translation errors. Recent Pieces of research develop a Neural Machine Translation (NMT) based on Statistical Machine Translation (SMT) whose performance is comparable to the performance of SMT, which is not satisfied anymore because of the huge number of errors found in the translated texts. Semantic web provides us with a mean to develop MTs whose performance can be satisfied compared with the performance of SMTs, according to the results provided by the researchers who developed the MT based on semantic web technologies.

Hutchins (1994) stated that the MT system might be designed to satisfy the following criteria: to deal with single words, to get restricted input text structure, and to have pre-edited input texts with any grammatical set, and without caring about the ambiguity of words or any other operators during translation. Ambiguity could be either nouns, verbs, or adjectives (Hutchins, 1994), but this depends on the properties of the language itself. The MT also was categorized by Hutchins (1994) into a large number of support groups (MT with machine assistance, MT with human aid, translation with computer support). Edited input or production version set of target words (deaf or multi-lingual), approaches to interpretation (immediate interpreting, interlingual or transfer-based) and evaluation of the syntactic composition (predictive assessment, sentence composition or grammar dependence). They also pointed to the importance of semantic MT, but it was just a survey, and there were no implementations for their vision (Hutchins, 1994). Chan et al. (2007), integrated a state-of-the-art Word Sense Disambiguation system into a state-of-the-art hierarchical phrase based MT, "Hiero". In contrast, they demonstrate only one way for the integration without introducing any rules that compete against existing standards (Gupta, Joshi, Mathur, 2013). Gupta et al. (2013) introduced 16 features that were extracted from the input sentence and their translation. Then showed a quality score based on Bayesian inference produced from their sample training data, but they did not develop a new English Hindi MT, and their work was just an analysis study for an existing MT (Gupta et al., 2013).

3.1 Statistical Machine Translation (SMT)

Statistical machine translation (SMT) is a term that refers to a group of MT systems that are developed using the machine learning methods. It is the most extensively studied type of MTs according to Srivastava et al (2016) who depicts a typical model for the SMT workflow (Srivastava, Rehm, Sasaki, 2017). The first SMT systems were developed less than three decades ago (Lopez, 2007). SMT is still the core of the new MT systems either those that are SMT in themselves, NMTs or MTs which are based on the semantic web. IBM in 1990 used A statistical method for machine transposition is developed alongside Bayes (Brown et al. 1990). López et al. (2007) claimed the SMT task is to take in the source language a sequence of words, toks or symbols that may be named A and transform into a sequence of terms, toks or symbols in the target language, which may be referred to as B (López, 2007). Lopez (2007) also classified the SMT, as shown in Fig. 1, according to the translation equivalence to:

Finite-State Transducer Models: extend two common finite-state automata sets (FSAs) of labels (the input and the output sets) and composed by making the output of one input to the next. Therefore, it needs to record the sentence before sharing the final translation (Lopez, 2007). It is also widely used in speech recognition. It was used as an efficient machine in the sequential string to-string and the sequential string-to weight transducers. The minimization and defeminisation algorithms for string-to weight transducers were used to solve some of the theoretical issues against the machine performance (Mohri, 1997). This type of models has two other different kinds (Lopez, 2007): o Word-Based Models: translates the source language word by word or more in the target language. When the user wants to translate a sentence, the most used meaning of a word will be used to translate that word always in the sentences (Lopez, 2007). It can also be used as a word segmentation as it was proposed by Sun with the Chinese language (Sun, 2010). It also can be used to align words that can be utilized in the phrase-based models while Machine Translation statistically (Lopez & Resnik., 2006).

Phrase-Based Models: translates the source language phrase by just a phrase in the target language. When the user wants to translate a sentence, that translation of the words will be shown in a sentence in the target language (Lopez, 2007). Callison-Burch et al. (2005), has rescaled this model to make it correspond with the longer phrases and the larger corpora. It is a way to solve the idioms problem. Besides, it is a widely used model because the error rate is small and it is fast in comparison with all other machine translations (Callison-Burch et al., 2005). However, this statistical or learning method is not applicable to identify the best meaning semantically because it replaces a phrase by another phrase in a different language. If there's a phrase that has a different meaning, the replacement mechanism will choose the most frequently used translation every time.

Synchronous Context-Free Grammar Models: Context free Grammar Generalization (CFG) may be used by syntactic parsing in transcription of one or both objects (Lopez, 2007). It can be used with the phrase-based model to develop the hierarchal phrase-based model which is used while building the Neural Machine Translation (NMT) (Chiang, 2005). It also can be used with the tree transducers promise to enhance the quality of the output of the SMTs where the complexity is exponential due to arbitrary reorderings between the two languages (Huang, Zhang, Gildea, & Knight, 2009). Decoding can be more flexible to synchronous context-free grammars by enhancing the unary rules by binarizing rules and phrases (Chung, Fang, & Gildea, 2011). This type of models has many other different types (Lopez, 2007):

- Bracketing Grammars: strict the number of rearrangements described by the synchronous context-free grammar model. However, the model sometimes cannot express the reordering relationship [33]. It is also used to induce the structure of the analogous translation grammar in the German-English alignment. It is done to disambiguate the meaning and to correct the bracketing errors (Zhang, Liu, Li, Zhou, & Zong, 2014, Carl, 2003).
- Syntax-Based Grammars: this is used in noisy-channel machine translation (Charniak, Knight, & Yamada, 2003) to easily merge the knowledge based on the natural language grammar rules. The needed result is to reorder the observations of the phrasal cohesion constraint imposed by exact linguistic syntax (Lopez, 2007).
- Hierarchical Phrase-Base Translation: as word-based designs, only word-to-word transformation is possible, and then a reordering choice is needed for each text (Lopez, 2007) which may trigger language errors, which contribute some vowels to some phrases in order to achieve the significance of the phrase. Every choice gives rise to more mistake opportunities. This model does not specify explicit term alignments, but it suggests that there may be an inner balance between phrases which appear in the same grammar law (Lopez, 2007). It is officially synchronous, but without any syntactic annotations, writing in conjunction is not synchronous (Chiang 2005).
- Syntactic paragraph-based models: Multi-word interpretation guidelines are included in phrase-based models. They are adorned by language text pieces that limit the feasible rearrangement of sentences (Lopez, 2007). Using a large n-gram language models is ideally the essential part in the best performance of the Machine Translation Systems (Brants, Popat, Xu, Och, & Dean, 2007). Incrementing the syntactic parsing into phrase-based translation will re-make the role of the language model as a mechanism for encouraging syntactically fluent translations (Schwartz, Callison-Burch, Schuler, & Wu, 2011).
- Alternative Linguistic Models (synchronous dependency grammar): Takes into consideration the rooted trees, which describe the phrase framework, the sentence words. You can transform many grammar dependences into context-free, and vice versa as well as conversions can be generated (Lopez, 2007). Dependency structures are naturally lexicalized as each node is one word. In contrast with phrasal structures (treebank style trees) which have two node types: lexical item terminals, and nonterminals that store word orders and phrasal scopes. A synchronous derivation process cannot deal with two types of cross-language mappings: crossing-dependencies and broken dependencies. A synchronous derivation process for the two syntactic structures of both languages proposes the level of cross-lingual isomorphism between the two trees of those structures (Ding, 2005).

Other Models of Translational Equivalence (tree-adjoining grammar): Linear Contenxt-Free Rewriting Systems (LCFRS) is one of the broad category of formalisms. These formalisms can analyze in linear moment a limited number of context-sensitive words. Theory remains a great deal of studies in this field (Lopez, 2007).

The statistical method was introduced as a new approach for MT in 1949 (Weaver, 1949). Statistical MT was stated by IBM researchers in 1986 (Brown et al., 1990). They thought that MT is as old as the first generation of the computers. They also said that the translation must depend on many factors and the essential one was the whole original text itself. In contrast, they treated the words without recognizing the connection between words or even recognizing the sentence structure. Koehn et al. (2003) indicated that studying sentences of more than 3 words and highly precise word-level matching sentences have no significant effect on efficiency. Learning only degrades the efficiency of their schemes by syntactically driven sentences.

3.2 Machine Translation using Semantic Web (SWMT)

A semantic web restructures the massive amount of data that is accessible and understandable to both humans and machines on the internet in a way that is similar to that of the human mind. It would be like training the internet to understand the context that is close to whatever word or phrase being searched through tags, the searcher attaches to the subject. The semantic web would serve as a way that helps computers think like the way humans do while still allowing the humans to understand those same files (Nandini, 2014). Using the semantic web, machines have the ability to deduce new facts from existing ones according to the available data. Besides, the semantic web enables computers to store and retrieve information and also produces new information. The ontology is the way that computers can understand the information they are carrying (Nandini, 2014). Moussallem et al. (2016), as shown in Fig. 2, classifies the MT according to their architecture, methodology, problem space addressed, and performance to:

Rule-Based Machine Translation (RBMT): To scan the input, create an intermediate linguistic representation; generate a text based on a morphological, syntactical and semantic map between the two words in their destination language.

- Direct approaches: In which phrases, without contemplating any differences in word significance, are converted one by one. This contributes to major mistake levels.
- Transfer-based strategies: Analysis of source language phrase composition, change to an intermediate framework, then generation of inner depiction on the basis of target language linguistic rules. Therefore, it utilizes three dictionaries: two monolingual dictionaries and a bilingual source-to-target vocabulary.
- Interlingua-based approaches: extracting the entry code independently and achieving a greater degree of precision than similar language techniques.

Corpus-Based Machine Translation: commonly use a large dataset containing the previous translations or a large set of examples to tackle the translation task. A bilingual corpus is essential for it.

- Statistical approaches: This is a model centered on this sort of MT. The MT is used to calculate and use a statistical translation model from the original corpus using computer training methods, including monitored, unattended and semi-supervised methods. In order to attain excellent precision, SMT needs big bilingual companies. The performance of the text therefore relies immediately on the chosen input corpus.
- Example-Based approaches (EBMT): It utilizes two-language businesses, but stores various information. The workout imitates a fundamental natural translator learning method and fills out a translation memory system database. The Task EBMT comprises of decomposition of source text, comparing and selecting instances of interpretation, adapting and recombining targets (Chun et al., 2017).

Hybrid Machine Translation: a group of both the rule-based and the corpus-based MTs.

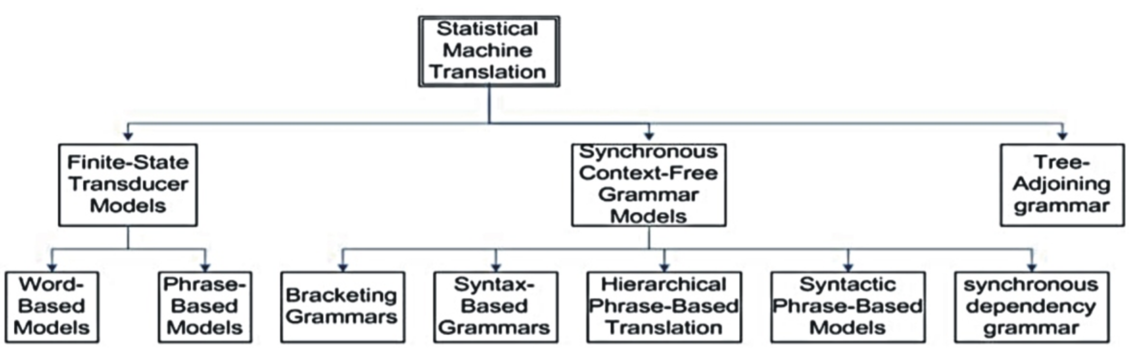

Figure 1: Statistical machine translations classification

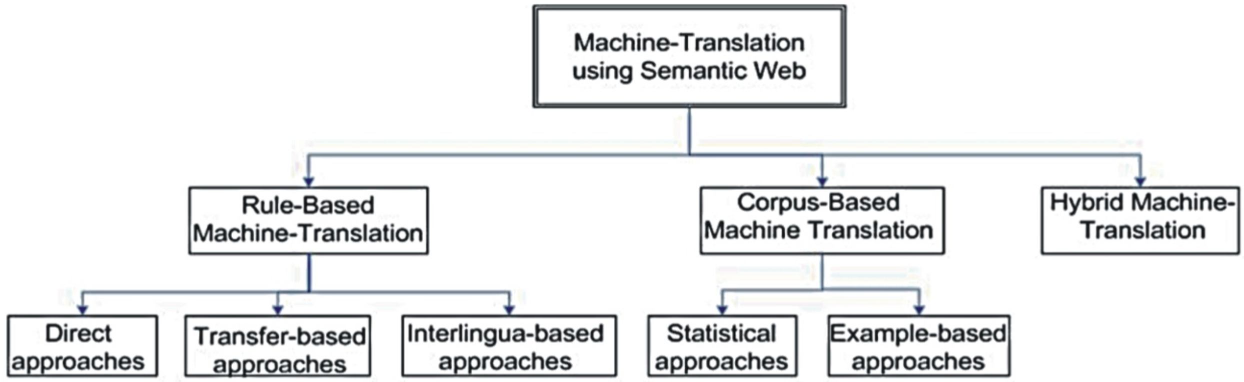

Figure 2: Classification of machine translations using semantic web

Shi et al. developed a world model for Chinese English MT using an Ontology-driven (Shi C, Wang, 2005). Seo, et al. (2009), presented a syntactic and semantic method for English-Korean MT using Ontology for web-based MT (AlAgha, 2015). Mousallem et al. (2015), solved the ambiguity in dialog conversation using an ontology-based MT (Moussallem et al., 2015). However, all of them can only serve for their proposed languages. They also mentioned that the instant translators such as Google, Bing, and more others are statistical translators and have no semantic processing in their works (Moussallem et al., 2015). Alagha et al. (2015), presented a domain-independent approach to translating Arabic NL queries to SPARQL by getting benefits from the linguistic analysis (Alagha et al., 2015). It was just to translate the Arabic questions to SPARQL, not to other human-languages. The systems developed in (AlAgha, 2015) taking questions expressed in Arabic and returns the answers drawn from an ontology-based knowledge base through the ontology file was not prepared as the MT system needs. While the translation is done to the SPARQL level only, it would not work as an MT between human-languages, and it would be like the previous works that didn't translate the Arabic question sentences to any other human-languages.

3.3 Neural Machine Translation (NMT)

Neural MT (NMT) is a newly engaged subject that works very badly with traditional sentence-based SMT techniques to the automatic translation workflow. The model NMT uses the Artificial Neural Network (ANN), instead of training the individual compounds of the MT independently, to jointly learn the model to maximize the translation effectiveness of both decoder and decoder by means of two phases of recurring neural network (RNN) (Cho et al., 2014). Dowmunts et al. (2016) made a comparison between the phrase based machine translation and the neural machine translation and found that they are equal, or the neural one gets somehow higher results than the phrase based. This is because the neural machine translation is used to remove the ambiguity using the n-gram technique and the results are to be sent to the phrase based machine translation to complete the translation. Besides, NMTs crucial goal is to build a single Neural Network that can be mutually adjusted to increase the translation performance (Bahdanau et al., 2014).

To the best of our knowledge, the Neural Machine Translation (NMT) is an encoder decoder architecture, according to Zhang et al. (2016), proposed by Blunsom and Cho et al. (2014) in respectively. NMT is an end-to-end MT that consists of two RNN (Zhang et al., 2016, Kalchbrenner, 2013), one is the encoder and the other is the decoder. The encoder network links each source sentence with a context vector representation according to a variable length of that sentence, and the decoder network produces the targeted translation word-by-word beginning from that context vector (Zhang et al., 2016). They proposed two approaches: handling source-side monolingual corpus in SMT and exploiting target side monolingual data in NMT. The source-side monolingual corpus is to use the source-side monolingual corpora to strengthen the encoder model of NMT, but they didn't fully explore it and they have no significant BLEU gains that are reported. Almahairi et al. (2016), develop their first NMT as a neural morphological analyzer based on a phrase-based statistical machine. The performance of that machine was comparable with the performance of the phrase-based statistical machine or slightly better. Dowmunt, et al. (2016) studied the neural machine translation based on the parallel corpus of six United Nations languages and phrase-based machine translation and has similar performance of Almahairi's. However, NMT cannot remedy a larger vocabulary due to the increase of the training and the decoding complexity proportionally with the number of target words (Long et al., 2017). Nematus was created by Sennrich et al (2017) as a toolkit that gives the NMT a strong focus on precision, usability and extensibility. Nematus was used for the preparation of top-performing applications for mutual WMT and IWSLT conversion activities and for the training of manufacturing environments technologies.

More examples of NMT include ByteNet and the Character-Based NMT. Kalchbrenner et al. (2016), developed an NMT to run in a linear time named ByteNet as a stack of two dilated convolutional neural networks, these two are the encoder and the decoder. Because the English and German strings encodings are sequences of characters, no clear division into words or morphemes is applied to the strings. The outputs are strings of characters in the target language (Kalchbrenner et al., 2016). Costajuss`a et al. (2016), proposes a character-based neural MT system which uses the character transformation representations which is embedded in a mixture of convolutional and highway layers. It is to increase the performance in many NLP tasks, including machine translation (Costajuss`a et al., 2016). OpenNMT is another example of NMT implementation that prioritizes efficiency and modularity (Kleiny et al., 2017).

Machine Translation using the deep learning (DLMT) seems like a branch of the NMT, but it differs from the NMT which is an end-to-end machine learning approach (Wu et al., 2016), by depending on pattern recognition, too (Schmidhuber, 2014). In contrast, Most of the existing NMT models are shallow, and this makes a performance gap between the single NMT model and the best conventional MT system (Zhou et al., 2016). Also, NMT systems are computationally expensive both in training and in translation inference. Therefore, Google uses the DLMT to bridge the gap between the human translation and the machine translation by developing a model that uses deep Long Short Term Memory (LSTM) networks with beside eight encoder and eight decoder layers via residual connections as well as attention connections from the decoder network to the encoder ones. They use an attention mechanism to connect the bottom layer of the decoder to the top layer of the encoder to improve parallelism, and as a result, to decrease training time. They employ low precision arithmetic through the inference computations to accelerate the final translation speed. Rare words were handled by dividing the words into a limited set of common sub-word units (“word pieces”) for the input and the output texts (Wu et al., 2016). However, the gap was just reduced but not completely avoided.

As a result of all those types of research, researchers are still trying to find a way that enables the machine to provide us an accurate translation. Because the SMT depends on pre-stored words or phrases translation (Lopez, 2007) and the NMT depends on the n-gram mechanism (Dowmunt, 2016), their translations sometimes are better than SWMT (Macketanz et al., 2017). However, they are all bad in metaphor translation because metaphor needs to be looked at from many sides (Macketanz et al., 2017) before providing the final translation. A comparison between the translation performance quality of the SMT, SWMT, and NMT shown in Table 1.

Table 1: Translation performance quality of the three types of MT

Phenomenon	SMT	SWMT	NMT
Homograph	Bad	Excellent	Bad
Terms	Good	Somehow bad	Somehow good
Idioms	Excellent	Bad	Excellent
Phrasal verbs	Very bad	Very good	Bad
Imperatives	Somehow good	Good	Good

4. Challenges of Machine Translation

It's not an easy job to translate a language. When converting one word into another, there are several difficulties to face. The following are the different difficulties:

4.1. Lexical ambiguity

Able to translate one language into another is one of the major problems. The phrases can have more than one significance in the source language as shown in Table 2. More than one significance can also be found in a set of phrases or a complete sentence as shown in table 3. The lexical ambiguity must be addressed if a computer is to comprehend and interpret accurately and it is a huge task to interpret the language accurately.

Table 2: Word with different meaning

English	Mizo
Read a **Book**	Lehkhabu chhiar
Book the flight ticket	Thlawhna ticket hauh rawh

Table 3: Sentence with more than one meaning

Sentence	Meaning
I saw **bats**	1.**Bats** are animals which can fly
	2. Multiple cricket **bats**

4.2. Differing word order

The structure of words or distinct term constructions may be distinct for two words. For instance, English is using a subject-verb-object (SVO) framework, whereas Mizo is using the Object-subject-verb (OSV) framework as shown in Table 4.

Table 4: Differing word order

Language	Sentence	Structure
English	I eat rice	SVO
Mizo	Chaw ka ei	OSV

4.3. Pronoun resolution

The pronoun replaces a sentence of a substantive or a substantive. The topic or the topic of the phrase as shown in Table 5 may be mentioned. Thus, the pronoun has to be addressed, which is a very difficult job for master interpretation, based on the phrase used.

It needs excellent understanding of language to construct a conversion system. However, the information acquired from the previous does not only require grammar, linguistics and language.

Table 5: Pronoun resolution

Sentence	**Pronoun (it) resolution**
The computer outputs the data, **it** is fast	**It** refers to the computer
The computer outputs the data, **it** is stored in ascii	**It** refers to the data

5. Conclusion

Translation, due to its significance in the field of traditional operations, study and discussion of buddies, and sometimes in politics and law, is the primary event in our everyday life. There are many elements, kinds and characteristics of translation. Translation has lately been a business, in particular with the fast acceleration in ICT, requiring the development of computers and instruments that make communication between the two parties easier. MT was one of the machines needed if in any communication process it wasn't the hidden agent. However, MT could not until this moment transcends the human translation and fortunately, the research has not stopped in this area and hopefully will continue to contribute in this field. There are three types of MT: SMT, SWMT, and NMT. Sometimes, some other tools are needed to be used during MT, to help analyze the text and then develop the texts in the target languages. These kinds of studies go on and will never halt until the MT can be superior or equivalent to human ones. We therefore say that there are still huge regions in which the system can be improved to use a completely automated version without the need for a professional assessment. In order to function in practice, the benefits of all the kinds of MT described in this study could require hybrid studies on the three methods.

References:

1. Abu-Tah,a, A.W., Alagha, I.M. (2015). An ontology based Arabic question answering system. Central library of Islamic University of Ghaza.
2. AlAgha, I. (2015). Using linguistic analysis to translate Arabic natural language queries to SPARQL. arXiv preprint arXiv.
3. AlAgha, I., Abu-Taha, A.W. (2015). AR2SPARQL: An Arabic Natural Language Interface for the Semantic Web. International Journal of Computer Applications, 1(6), 126-137.
4. Almahairi, A., Cho, K., Habash, N., & Courville, A. (2016). First Result on Arabic Neural Machine Translation. arXiv preprint arXiv, 16(6), 75-86..
5. Attia, M. (2006). An ambiguity-controlled morphological analyzer for modern standard arabic modelling finite state networks. Challenges of Arabic for NLP/MT Conference; London, UK.: The British Computer Society.
6. Ba-Alwi, F.M., Albared, M., & Al-Moslmi, T. (2017). Choosing the Optimal Segmentation Level for POS Tagging of the Quranic Arabic. British Journal of Applied Science & Technology, 19(1):10.
7. Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv ISO, 1(4), 9-27.
8. Bar-Hillel, Y. (1960). The present Status of Automatic Translation of Languages. Advances in Computers, 1:91-163.
9. Boucau, F. (2010). The European Translation Markets: Updated facts and figures: Brussels: EUATC.
10. Bowker, L. (2002). Computer-aided translation technology: A practical introduction. University of

Ottawa Press.

11. Brants, T., Popat, A.C., Xu, P., Och, F.J, Dean, J. (2007). Large language models in machine translation. Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning; 2007: CiteSeerx.
12. Brown, P.F, Cocke, J., Pietra, S.A.D., Pietra, V.J.D., Jelinek, F., Lafferty, J.D. (1990). A statistical approach to machine translation. Computational linguistics, 16(2):79- 85.
13. Callison-Burch, C., Bannard, C., Schroeder, J. (2005). Scaling phrase-based statistical machine translation to larger corpora and longer phrases. Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics.
14. Carl, M. (2003). Inducing Translation Grammars from Bracketed Alignments: Springer Netherlands, 339-361.
15. Chan, Y.S., Ng, H.T., & Chiang, D. (2007). Word sense disambiguation improves statistical machine translation. Annual Meeting Association for Computational Linguistics,3(3), 12-19.
16. Charniak, E., Knight, K., Yamada, K. (2003). Syntax-based Language Models for Statistical Machine Translation. Proceedings of MT Summit IX.
17. Chiang, D. (2005). A hierarchical phrase based model for statistical machine translation. Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics; Association for Computational Linguistics.
18. Cho, K., Merriënboer, B.V., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv.
19. Chua, C.C., Lim, T.Y., Soon, L.K., Tang, E.K., & Ranaivo-Malançon, B. (2017). Meaning preservation in example-based machine translation with structural semantics. Journal of Expert Systems with Applications.
20. Chung, T., Fang, L., Gildea, D. (2011). Issues concerning decoding with synchronous context-free grammar. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers; Association for Computational Linguistics.
21. Costa-Jussà, M.R., & Fonollosa, J.A.R. (2016). Character-based neural machine translation. arXiv preprint arXiv.
22. Dickins, J., Hervey, S., & Higgins, I. (2002). Thinking Arabic translation: A course in translation method: Arabic to English. Cambridge, UK: Routledge.
23. Ding, Y., & Palmer, M. (2005). Machine translation using probabilistic synchronous dependency insertion grammars proceedings of the 43rd Annual Meeting of the ACL; 2005: Association for Computational Linguistics.
24. Ditters, E. (2001). A formal grammar for the description of sentence structure in modern standard Arabic. EACL 2001 Workshop Proceedings on Arabic Language Processing: Status and Prospects
25. Group, T.W.B. (2004). The World Bank Translation Style Guide. Arabic Edition ed. Group TWB, editor. Washington, United States of America: The International Bank for Reconstruction.
26. Gupta, R., Joshi, N., & Mathur, I. (2013). Analyzing quality of English-Hindi machine translation engine outputs using Bayesian classification. International Journal of Artificial Intelligence and Applications, 4(4), 165-71.
27. Han, A.L.F, & Wong, D.F. (2016). Machine Translation Evaluation: A Survey. arXiv preprint arXiv, 1(2), 34-49.
28. Hatim, B., & Munday, J. (2004). Translation: An advanced resource book: Psychology Press.
29. Huang, L., Zhang, H., Gildea, D., & Knight K. (2009). Binarization of synchronous context-free grammars. Computational Linguistics, 35(4), 559-95.
30. Hutchins, J. (1994). Machine translation: History and general principles. The encyclopedia of

languages and linguistics, 5, 2322- 2332.

31. Hutchins, W.J., & Somers, H.L. (1992). An introduction to machine translation. London: Academic Press.
32. Indurkhya, N., & Damerau, F.J. (2010). Handbook of Natural Language Processing. Second Edition ed. R. Herbrich TG, editor. Microsoft Research Ltd.: Cambridge, UK.
33. Jakobson, R. (1959). On linguistic aspects of translation. On Translation, 3, 30-39.
34. Junczys-Dowmunt, M., Dwojak, T., & Hoang, H. (2016). Is neural machine translation ready for deployment? a case study on 30 translation directions. arXiv preprint arXiv.
35. Kalchbrenner, N., & Phil, B. (2013). Recurrent continuous translation models. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing.
36. Kalchbrenner, N., Espeholt, L., Simonyan, K., Oord, A.V.D., & Graves, A. (2016). Kavukcuoglu K. Neural Machine Translation in Linear Time. arXiv preprint arXiv.
37. Kleiny, G., Kim, Y., Deng, Y., Senellart, J., & Rush, A.M. (2017). OpenNMT: Open-source Toolkit for neural machine translation. arXiv preprint arXiv.
38. Koehn, P., Och, F.J., & Marcu, D. (2003). Statistical phrase-based translation. Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology; 2003: Association for Computational Linguistics.
39. Kuzenko, G.M. (2008). The World of Interpreting and Translating. Mykolayiv: MSHU. Peter Graves.
40. Long, Z., Mitsuhashi, T., Mitsuhashi, T., & Yamamoto, M. (2017). Translation of patent sentences with a large vocabulary of technical terms using neural machine translation. arXiv preprint arXiv.
41. Lopez, A. (2007). A survey of statistical machine translation (No. UMIACS-TR-2006-47). Maryland Univ College Park Dept of Computer Science.
42. Lopez, A., & Resnik, P. (2006). Word-based alignment, phrase-based translation: What's the link? Proc of AMTA.
43. Luo, Y., & Bu J.(2016). How valuable is information and communication technology? A study of emerging economy enterprises. Journal of World Business, 5(1), 200211.
44. Macketanz, V., Avramidis, E., Burchardt, A., Helcl, J., & Srivastava, A. (2017). Machine translation: Phrase-based, rule-based and neural approaches with linguistic evaluation. Cybernetics and Information Technologies, 17(2):16.
45. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J.R., Bethard, S., & McClosky, D. (2014). The Stanford Core NLP natural language processing toolkit. ACL (System Demonstrations).
46. Martin, J.H., & Jurafsky, D. (2000). Speech and language processing. International Edition, 710.
47. Mohri, M. (1997). Finite-state transducers in language and speech processing. Computational Linguistics, 23(2), 269- 311.
48. Moussallem, D., & Choren, R. (2015). Using Ontology Based Context in the Portuguese-English Translation of Homographs in Textual Dialogues. International Journal of Artificial Intelligence & Applications, 6(5):17- 33.
49. Moussallem, D., Ngomo, A.C.N., & Wauer, M. (2016). Machine Translation Using Semantic Web Technologies: A Survey. IOS Press, 1(5):1-22.
50. Nandini, D. (2014). Semantic Web and Ontology e-Books and from bookboon.com. Booknoon.com.
51. Newmark, P. (1988). A textbook of translation. New York: Prentice hall.
52. Othman, E., Shaalan, K., & Rafea, A. (2003). editors. A chart parser for analyzing modern standard Arabic sentence. Proceedings of the MT summit IX workshop on machine translation for Semitic languages: issues and approaches, USA.
53. Sawalha, M., Atwell, E., Abushariah, M.A. (2013). SALMA: standard Arabic language morphological analysis. Communications, Signal Processing, and their Applications (ICCSPA),

2013 1st International Conference on: IEEE.
54. Schmidhuber, J. (2014). Deep learning in neural networks: An overview. arXiv preprint arXiv.
55. Schwartz, L., Callison-Burch, C., Schuler, W., & Wu, S. (2011). Incremental Syntactic Language Models for Phrase-based Translation. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies; 2011: Association for Computational Linguistics.
56. Sennrichy, R., Firat, O., Cho, K., Birch, A., Haddow, B., & Hitschler, J. (2017). Nematus: A toolkit for neural machine translation. arXiv preprint arXiv.
57. Seo, E., Il-Sun, S., Su-Kyung, K., & Ho-Jin, C. (2009). Syntactic and semantic English-Korean machine translation using ontology. Advanced Communication Technology, 2009 ICACT 2009 11th International Conference on; 15-18 Feb. 2009: IEEE.
58. Shi, C., Wang, H. (2005). Research on ontology-driven Chinese-English machine translation. Natural Language Processing and Knowledge Engineering, 2005 IEEE NLP-KE '05 Proceedings of 2005 IEEE International Conference on; 2005, 30 Oct.-1 Nov. 2005; Wuhan, China: IEEE.
59. Srivastava, A., Rehm, G., & Sasaki, F. (2017). Improving machine translation through linked data. The Prague Bulletin of Mathematical Linguistics, 108:12.
60. Sun, W. (2010). Word-based and character based word segmentation models: Comparison and combination. Proceedings of the 23rd International Conference on Computational Linguistics: Posters 2010: Association for Computational Linguistics.
61. Taylor, C. (2009). The Routledge Handbook of Translation Studies. Revised Edition ed. Munday J, editor: Routledge.
62. Venuti, L. (2008). The translator's invisibility: A history of translation, Routledge.
63. Weaver, W. (1949). Translation. Machine translation of languages, 14,15-23.
64. Wołk, K., & Marasek, K. (2015). Neural-based machine translation for medical text domain. Based on European Medicines Agency leaflet texts. Procedia Computer Science.
65. Wu, Y., Schuster, M., Chen, Z., Le, Q.V., Norouzi, M., & Macherey, W. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv.
66. Zaghal, R., & Sbeih, A.H. (2012). Arabic morphological analyzer with text to voice. In Communications and Information Technology (ICCIT), 2012 International Conference on; 26-28 June 2012; Hammamet, Tunisia: IEEE. 2012;48-52.
67. Zhang, J., & Zong, C. (2016). Exploiting source-side monolingual data in neural machine translation. In Proceedings of EMNLP ISO 690.
68. Zhang, J., Liu, S., Li, M., Zhou, M., & Zong, C. (2014). Mind the Gap: Machine Translation by Minimizing the Semantic Gap in Embedding Space. Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence; 2014; China: Zong Publications.
69. Zhou, J., Cao, Y., Wang, X., Li, P., & Xu, W. (2016). Deep recurrent models with fast-forward connections for neural machine translation. arXiv preprint arXiv.

Note: This research is funded by the Deanship of Scientific Research at Zarqa University-Jordan