Translation of Idiomatic Expressions from Arabic into English using AI (Artificial Intelligence)

[1]Shahab Ahmad Almaaytah

[1]*Applied College in Abqiaq King Faisal University Assistant Professor , Human science and languages - Saudi Arabia (KSA) - salmaaytah@kfu.edu.sa*

Abstract

Translation entails conversion of words from SL (Source Language) to TL (Target Language) that includes form, meaning, context, culture and expression. Idiom is a multiword expression and its meaning cannot be translated verbatim because idioms are fixed non-compositional expressions that belong to the intersection between language and culture. This makes the translation tricky as there can be a number of options available as meanings for every idiom that is translated into the target language. The precise translation of the idiom is possible if there is an equivalent in the target language but when the equivalent idiom is not available in the target language it has to be paraphrased in such a manner using certain strategies so that the meaning is clear while at the same time the translation should not affect the cultural aspects that play an important role. Therefore, translation of idioms from the source language to the target language must take into consideration the cultures of both languages. In this paper, suggestions are given on how Arabic idioms can be translated into English.

Keywords: English, Arabic, Translation, Idioms, SL, TL, Strategy, Culture.

INTRODUCTION

Idioms are expressions used in natural language to make language more colourful, picturesque and dramatic, and this is the natural way in which people speak the language. Individual words in idioms do not express the meaning that the entire idiom conveys. Therefore, the translation of an idiom cannot follow the principles or strategies used in translating a simple sentence such as 'هذا قلم' which means 'this is a pen'.

Idioms are figurative expressions that are made up of at least two words which cannot be understood literally and which function as a unit semantically1. Though idioms are used both in spoken as well as written languages, they are typically used in spoken language2. There are many different categories of idioms3 and Arabic conversations are full of expressions such as4 'عم استنى بفارغ الصبر' and it means 'To be waiting impatiently or nervously for someone or something'. As will يطويها النسيان /يبتلعها العدم / as and it means 'fade into the sunset.

When an idiom is translated perfectly, it does not look like a translation but a smooth transition of words from the SL to TL due to the style and voice that carries the meaning that is close to the SL but very natural in TL. In this paper, the expressions and meanings of some Arabic idioms that have an English equivalent and a few that do not have an equivalent in English are taken as examples to show how translation gets affected and what strategies can be used to get the nearest English equivalent for the translation of an Arabic Idiom into English.

Literature review

Mr. Marwan Akeel and Ravi Mishra in their article entitled 'ANN and Rule Based Method for English to Arabic Machine Translation' (The International Arab Journal of Information Technology, Vol.11, No.4, July 2014) discussed World Wide Web, globalization and international business bring the world together. In order to have smooth communication, the need of breaking the language barriers appears. Here comes the study of machine translation systems, which is the study of designing systems that translate text from one human language into another language with or without minimal human intervention. Machine translation is an automated process, in which translation job is done by the computer software. Machine translation is an application of computer linguistics. Computer linguistics is an interdisciplinary field of computer science and requires language and computer experts. Translation as an art of rendering a work of one language into another is as old as written literature. Computer technology has been applied to technical translation to improve one or both of the factors like, Speed: Translation by or with the aid of machines can be faster than manual translation and Cost: Computer aids to translation can reduce the cost per word of a translation.

If MT researchers are able to develop a perfect multilingual machine translation system, people with different languages can share ideas and information worldwide on every topic as business, economic, educational, political, socio-cultural, etc. The purpose of a translation process is that meaning of the text being translated should not change. There are many different machine translation systems available online or desktop systems.

Mansoor Al-A'ali in his article entitled 'Pre-editing and Recursive-Phrase Composites for a Better English-to-Arabic Machine Translation' (Journal of Computer Science 3 (6): 410-418, 2007 ISSN 1549-3636 © 2007 Science Publications) explained Machine Translation techniques have recently achieved some success but in spite of this success, MT still has many hurdles to overcome. Recent research

contributed to the basic principles of machine translation systems by using phrase translation. Most phrase-based machine translation systems rarely use linguistic knowledge of the structure of the languages involved. At the same time, the cost for using these techniques is high relative to the small improvements gained in performance. Some researchers attempted to employ parsers for tree-to-tree and tree-to-string alignment, respectively. Various experiments have been attempted on the effect of varying amounts of morphosyntactic information, including the techniques used. Linguistic knowledge from NLP tools can be used effectively with relatively small training data and in a limited domain. The problem is how reliable and accurate the knowledge is, and how to employ it. A major challenge in machine translation is how to build phrase translation which takes into consideration all aspects of the translation process such as semantics and context. There are normally three different methods to build phrase translation probabilities: learning phrase alignments from a word-aligned corpus, learning syntactically motivated phrase pairs from a word aligned and parse tree annotated corpus, and learning a joint phrase model. Phrase pairs consistent with a word alignment have so far yielded the highest performance. Syntactic phrase pairs can be restricted by both the word-alignments and the source and target parse trees. The authors examined the effect of using tags and syntactic chunks, and treebank-based syntactic parses of source and target sentences within an n-best re-scoring framework based on a log-linear model. Chinese sequences were projected onto the English words using the word alignment. Relative positions were indicated for each Chinese tag. Then a trigram language model was built on these projected Chinese tags and positions. The projected information was one of the most beneficial syntactic features used. Among the syntactic features evaluated, the simple Markov model achieved the most significant performance improvement.

Dimitra Anastasiou in her Ph.D. thesis entitled 'Idiom treatment experiments in machine translation' in 2010, embarked upon idiomatic

expressions pose a particular challenge for the today's Machine Translation systems, because their translation mostly does not result literally, but logically. The present thesis shows, how with the help of a corpus, and morphosyntactic rules, such idiomatic expressions can be recognized and finally correctly translated. The work leads the reader in the first chapter generally to the field of Machine Translation and following that, it focuses on the special field of Example-based Machine Translation. Next, an important part of the doctoral thesis dissertation is devoted to the theory of idiomatic expressions. The practical part of the thesis describes how the hybrid Example-based Machine Translation system METIS-II, with the help of morphosyntactic rules, is able to correctly process certain idiomatic expressions and finally, to translate them. The following chapter deals with the function of the transfer system CAT2 and its handling of the idiomatic expressions. The last part of the thesis includes the evaluation of three commercial systems, namely SYSTRAN, T1 Langenscheidt, and Power Translator Pro, with respect to continuous and discontinuous idiomatic expressions. For this, both small corpora and a part of the extensive corpus Europarl and the Digital Lexicon of the German Language in 20th century were processed, firstly manually and then automatically. The dissertation concludes with results from this evaluation.

Arwa Abdulaziz Assaf Al-assaf in his Masters degree dissertation entitled 'Translating Idioms from English into Arabic: Appointment with Death as a Case study' said, in recent studies, more focus has been targeted towards descriptive and comparative translation studies. Translation has customarily been seen as an interlingual correspondence though some translation scholars, as Hans J. Vermeer, are energetically against the origination of translation principally as a culturally diverse exchange or intercultural correspondence (Bassnett and Lefevere, 1990).

The spread of the human race through the course of time came along with a variety of languages and dialects that created a profound hole of disconnection among the languages users because of their insufficiency in translating the message of source language to their own language. Progressively, this need of having the capacity to understand other languages set the way to the rise of a new discipline, concerned with expressing a language in different words of another language which is known mostly by translation.

Translation scholars as well as etymologists all concur that there is a nearby connection in the middle of language and society. Vermeer (1992) considers dialect to be a characteristic piece of society, whose perspective is shared by Mary Snell-Hornby especially when culture is characterized as "a totality of information, capability and discernment" (Nord, 1997, p.23). Lyons (1968), an English etymologist, as well, considers the language of a specific culture as an indispensable piece of its way of life; he also holds that the lexical refinements drawn by every language will have a tendency to mirror the socially vital components of items, establishments and exercises in the general public in which the language works.

Similarly mindful of the entwined state of language and culture, Nida (1993) trusts that since culture is summed up as "the totality of convictions and practices of a general public, nothing is more deliberately critical than the language through which its convictions are communicated and passed on and by which most association of its individuals happens."(p.105). Munday (2001,p.127) reports that "Bassnett and lefevere (1990) go beyond language and focus on the interaction between translation and culture, on the way which culture impacts and constrains translation and on the larger issues of context, history and convention". Regardless of whether those scholars view language as a component of culture or not, they all recognize that language and culture are firmly associated or related. It is, consequently, not astounding that language and culture are here and there discussed as "languaculture"- a solitary substance (Nord, 1997,p.25). Bassnett (1980) states that Since language and culture are interrelated, and translation is considered as an interlingua communication and additionally a procedure of social exchange, translators are quite often required to be both bilingually and culturally

skillful. The translator, as an extraordinary kind of reader and his very own insider society, cannot generally be a passkey to opening all the social locks. He cannot totally rise above the limits of his own social legacy into the universe of the other. He himself, some of the time, may be excessively bolted into his own social state of mind to have the capacity to share different societies. Consequently, the level of trouble and the nature of translation have more to do with culture than with language itself. "Translation is a sort of action which definitely includes no less than two dialects and two cultures customs."(Toury1978:200).

As this announcement suggests, translator are forever confronted with the issue of how to treat the social angles verifiable in a source content (ST) and of discovering the most suitable procedure of effectively passing on these viewpoints in the objective dialects (TL). Further he said the intricacy of issues which can be experienced by a translator in translation of idioms from English into Arabic. Bassnett states that "the translation of idioms takes us a stage further in considering the topic of which meaning and translation, for idioms, similar to puns, are culture bound"(p.32). At the point when two dialects have compared informal expression that renders the idea of prevarication, thus during the time spent Interlingua translation of one idiom substituted for another. That substitution is made neither on the premise of the semantic components in the expression, nor on the premise of a relating or comparative picture contained in the expression, yet in the phrase. The SL expression is supplanted by a TL expression that fills the same need in the TL society, and the procedure includes the substitution of SL sign for TL sign.

Definition of Idioms

Many scholars have defined idioms in their own words and in this section a few significant definitions have been discussed. One of the criteria used for defining an idiom is that it is an expression whose meaning is unpredictable based on the individual words used in the

idiom. Even a two word phrase such as 'stand up' can be considered as an idiom as per this definition and so Pulman 1993 states that 'multi-wordedness' of a phrase alone is not enough to consider the phrase as an idiom though it is an essential aspect of an idiom.

To define an expression as an idiom the characteristics accepted by scholars have changed over the years6. Some scholars like Hockett, 1964; Katz and Postal, 1963; Makkai, 1972 use the formal characteristics to define an idiom and this includes the grammatical form such as morphemes which are single words whereas 'lexical idioms' and 'phrasal idioms' are two types of idioms defined by Katz and Postal who argue that the literal meaning and the idiomatic meanings are completely different for these words or multiword expressions. Makkai (1972) defined idioms as expressions that have at least two words but this cannot be accepted as many simple expressions or phrases will also have to be considered as idioms.

An approach called the generative transformational approach was used by Fraser (1970) and Wienreich (1969) in order to study about idioms. According to Wienreich (1969) idioms are phrases that have the same meanings literally and figuratively, which restricts many idioms from being considered as idioms. Fraser (1990) states that idioms can have variations and developed a scale which shows the degree up to which an idiom can be pressed to get a different meaning. Fernando (1996) uses both the degrees of variance as well as the level of transparency of an idiom in order to define it. Moreover, both Fraser and Fernando define idioms based on the characteristics of the idiom.

Idioms may be considered cultural and as involving trope. According to Oxford Dictionary, "an idiom is a group of words established by usage as having a meaning not deducible from those of the individual words". As idioms are fixed expressions they cannot be divided into their individual elements for getting the actual meaning. . Furthermore, an idiom has two important functions namely pragmatic and referential.

The pragmatic function helps the receiver to show an interest as this appeals to the senses whereas the referential function is described as "a mental process or state, a concept, a person, an object, a quality or an action more comprehensively and concisely than is possible in literal or physical language" Newmark (1988).

Idioms have also been defined by various scholars. Keeping the objective of the current study in mind, an idiom can be considered as an expression that has at least two words and the meaning of the idiom cannot be deduced or predicted from the meanings of the individual words that make up the idiom.

Classification of Idioms

Different scholars have classified idioms based on various factors that they have studied in their research. Moon (1998) states that the term 'idiom' is used as a common expression for a variety of multiword expressions whether the expression is opaque or transparent. Idioms have been classified into three types namely Pure-idioms, Semi-idioms and Literal-idioms by Fernando (1996) based on the transparency of meaning. Pure-Idioms do not change their meaning and are opaque in nature. An example of a pure idiom is: to kick the bucket which is considered to be semantically opaque and lexically fixed expression. In a semi-literal idiom at least there is one element that is non-literal in nature and these idioms according to Fernando are not fully opaque but are partially opaque. A typical example of a semi-literal idiom is to cost an arm and a leg with the verb element, to cost, preserving its literal sense when used in this idiomatic expression .The last type in Fernando's classification is 'Literal-idioms' which are less complex semantically and the variations are restricted to a certain degree. An example of a literal idiom is : I felt numb and had to shake a leg to stand up .

Strategies used by interpreters to translate idioms

Human translators used certain strategies to translate idioms from SL to TL and these strategies have been studied by Baker (1992) and categorized into four major types.

a.   Using a similar idiom in TL

b.   According to Baker (1992) ,if a similar idiom in TL is used in translation, then the meaning of the text does not get altered and so the reader does not find any problem in the translated version of the original text. For example يعطي الضوء الاخضر in Arabic means 'To give a green signal' or اصطاد في الماء العكر in Arabic means to fish in troubled waters. There are many such similar idioms in English that have the same meaning as in Arabic and these idioms can be used in translation of Arabic idioms to English. If both the meaning and form of the idioms in the source language are not changed during translation then the sense of the text is not lost and so the reader finds the easy flow of target language, not interrupting the understanding of the whole text

c.   Idioms with Similar meaning but in different forms

Some idioms that have matching English idioms but the way they are expressed is not the same. Translators sometimes use these idioms in order to convey the meaning in the target language so that the translation does not deviate much from the contents of the source language. For example إنك لا تجني من الشوك العنب can be translated in English as 'one can't get blood from stone'. The literal meaning is 'You do not reap the grape of thorns'

Similarly, رجع بخفي حُنينٍ can be translated into English as 'empty handed' but just as in the example above the way it is expressed in Arabic is not the same as in English. In this expression (رج) means 'came back' and (بحفي) means 'it's a name used for a kind of shoes old name'.

d.   Borrowing the idiom from the source language

Like loan words there are phrases and idioms that are also borrowed from Arabic into English. One of the most commonly used idioms that was borrowed from Arabic is "The last straw". The original version was "The straw that broke the camel's back" which

means that a person's patience has come to an end. The original phrase in Arabic is "alqasshaalatheeqassamatdhahra al baeer" and its literal meaning is "the straw that broke the back of the camel."

Another common idiom in English is 'a camel's nose' which comes from an old Arab proverb that says when an Arab allowed his camel to just get its nose into a tent it ended up with the Arab moving out of the tent because the camel occupied the entire tent over a short period. Thus, something that looks as a simple act can turn out to be an adverse action in a short span of time.

e.   Omitting Idioms

Baker's final strategy is to completely avoid an idiom that does not have a match in the target language. There are many idioms in Arabic that do not have an equivalent in English and also cannot be translated into English due to many factors including the fact that they are culture-specific.

Some examples from spoken Arabic are from Barakbits, 2017:

1)   Maqtoo' min shajara مقطوع من شجرة – Literal meaning: 'cut from a tree'. When a person lacks a family or is an orphan this expression is used to explain that he has no family tree to be identified with.

2)   Eid wahda matsa'afsh يد واحدة ماتسقفش - Literal meaning: 'one hand doesn't clap'. The figurative meaning is that a problem is caused by more than one person or it can also be used to express that everybody in a team has to work for anything positive to materialize. This essentially means that if something is to work, cooperation from all parties is necessary, usually said to encourage teamwork.

Similarly, in Quran there are many idioms that cannot be translated perfectly as classical Arabic is a language that has words with a variety of meanings and each meaning has a different level of intensity. The spiritual sense that is to be translated is not easy unless the translator is aware of both language and religion apart from the differences in both the cultures.

Machine Translation

The use of computers to translate texts from Source Language to Target language is known as Machine translation[7] and translation of idioms is not easy[8]. Machine translation has advanced technologically since 1950[9] and statistical machine translation that had been basically developed by IBM has undergone a lot of sophistication in order to get results that are as perfect as possible in the target language. In the later models of SMT the translations have been conducted on phrases or on a string of words, which are available in bilingual corpora and some universities have been supported and funded by Google Corporation[10]. This technology needs processing power because each day free translation websites get a large number of translation requests; Google gets more than a billion translation requests[11] which is a very large volume.

Translation memories were used to store large database of translated phrases so that the machine could easily translate large amounts of texts faster than in the earlier stages when translation of source text to target text consumed too much time[12]. Translation memories are used by various companies that are global in nature as these memories are cost-effective and also consistent[13]. These are used in the neural network in order to increase the efficiency of translation[14]. Statistical machine translation produces many different translations for each sentence giving the user more choices to choose from which is a paradigm shift from the previous rule based translation methods[15].

1.   Koehn and Senellart, 2010

2.   Adelnia and Dastjerdi, 2011

3.   Hutchins, 2005

4.   Hutchins, 2005

5.   Specia, 2012

6.   Mohammed, 2011

7.   Seljan and Pavuna, 2006

8.   Sukhbaatar, Weston, and Fergus; 2015

9.   Vashee, 2013

Artificial Intelligence and Translation

Since 2014, there have been further changes in translation technology. Earlier Machine translation followed a system called translation memories that enhanced the quality of output in machine translation but since 2014, the concept of neural translation has been designed that can build a network of artificial neural networks, which works on a very high volume of translation memories16. In this system, a single network will be built to give the best possible translation which consists of a number of encoders and decoders. These encoders-decoders produce translation of the source language into a fixed-length sentence in the target language. Statistical models are based on the biological models where layers of artificial neurons are created17   therefore, this system uses lesser number of components as compared to the earlier statistical machine translation which helps in speeding up the translation process.

Discussion

The current study reveals that Arabic as an idiomatic language has various categories of idioms such as: colloquialisms, proverbs, slang, allusions and phrasal verbs. In addition, there are some strategies that are used by interpreters to translate idioms namely; using a similar idiom in TL, idioms with similar meaning but in different forms, borrowing the idiom from the source language and omitting idioms. Finally, Artificial Intelligence plays a crucial role while translating idioms from Arabic to English especially in the era of technology and mass media.

Conclusion

A language's beauty is in its ability to communicate any expression and if the language has idiomatic expressions then the quality of the language is considered to be very high. Arabic is a rich language and the idiomatic expressions in Arabic make the language very attractive. Most people who speak the language use idioms naturally without even thinking about the idioms they are using from the time they learn to speak the language. Idioms are a natural part of any language, spoken or written.

In the field of machine translation major changes have taken place from the time of inception in the 1950s till now but there are certain difficulties in translation even in this decade especially the translation of idioms form Arabic to English. In order to get the best translation I propose a few suggestions. Firstly the machine translation system must be able to segregate the phrases from idioms. Once the idioms are segregated the next process by the software is to find out the idiom's classification and give a suitable translation that is taken out from database available in the translation memory.

Artificial Intelligence can be used to get both the naturalness as well as giving the correct meaning of the text. If the Arabic idiom does not have a corresponding English idiom then the machine should be used to output the paraphrased text so that the final output does not have any errors. This would help in getting the naturalness and flow of language in the translated text that is as accurate as possible.

Reference

[1] Adelnia, A., & Dastjerdi, H. V. (2011). Translation of idioms: a hard task for the translator. Theory and Practice in Language Studies, 1(7), 879-883.
[2] Allan Keith 2001. Natural Language Semantics. Oxford: Blackwell.
[3] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

[4] Beekman, J., & Callow, J. (1974). Translating the word of God (pp. 74-4950). Michigan, USA: Zondervan Publishing House.

[5] Hutchins, J. (2005). The history of machine translation in a nutshell. Retrieved December, 20, 2009.

[6] Koehn, P., & Senellart, J. (2010, November). Convergence of translation memory and statistical machine translation. In Proceedings of AMTA Workshop on MT Research and the Translation Industry (pp. 21-31).

[7] Leah, C. (2014). Idioms-Grammaticality and Figurativeness. University of Oradea.

[8] Mohammed, T. A. S. (2011). A taxonomy of problems in Aarabic-Eenglish Translation: a systemic functional Linguistics approach Tawffeek abdou.

[9] Reijonen, H. (2006). General characteristics and origins of English idioms with a proper name constituent: a study based on their etymology as available in the typical compilation.

[10] Specia, L. (2012). Statistical Machine Translation. Emerging Applications of Natural Language Processing: Concepts and New Research: Concepts and New Research, 74(8), 15-19.

[11] Seljan, S., & Pavuna, D. (2006, September). Translation Memory Database in the Translation Process. In Proceedings of the 17th International Conference on Information and Intelligent Systems IIS 2006 (pp. 327-332). Croatia, Varaždin: FOI, 2006a.

[12] Sukhbaatar, S., Weston, J., & Fergus, R. (2015). End-to-end memory networks. In Advances in neural information processing systems (pp. 2440-2448).

[13] Tu, Z., Liu, Y., Shang, L., Liu, X., & Li, H. (2017, February). Neural Machine Translation with Reconstruction. In AAAI (pp. 3097-3103).

[14] Vashee, K. (2013). Understanding the economics of machine translation. Translation Spaces, 2(1), 125-149.