International Journal of Multidisciplinary Thought,
CD-ROM. ISSN: 2156-6992 :: 3(4):1–16 (2013)

DIFFICULTIES IN MACHINE TRANSLATING TEXT FROM ENGLISH INTO ARABIC

Amna Mohamed Benzayed,

Faculty of Arts, University of Tripoli, Libya

Machine Translation (MT) is the use of computers in translating text from one natural language to another. This paper begins by an overview on machine translation (MT). Then, it presents an alternative means that could help expedite translation, which has been tested and proven successful to a certain extent, is machine translation (MT) systems. Systems such as SYSTRAN, ARIANE, METEO, and a few others are examples of systems that have been developed and utilised quite successfully. This study shed the light on some problem of machine translation.

INTRODUCTION AND OVERVIEW:

In the recent years, the need for more translations than human translators can produce has increased all over the world and led to the development of modern technology. The mechanisation of translation has been one of humanity's dreams which have become a reality in the twentieth century, in the form of computer programs capable of translating a wide variety of texts from one language into another. But there are no translating machines which produce a perfect translation in any other language without human intervention. Most translation in the world is not of texts which have literary and cultural status. The great majority of professional translators are employed to satisfy the huge and growing demand for translations of scientific and technical documents, commercial and business transactions, administrative memoranda, legal documentation, instruction manuals, agricultural medical text books, industrial patents, publicity leaflets and newspaper reports .

The use of a machine to translate one language to another was first suggested by Booth in the course of a conversation with Warren Weaver (1946). At that time the problem was simply an intellectual exercise directed at finding yet another use for the new high-speed digital calculators, which were just coming into existence. Little support was given to machine translation. Coupled with the lack of calculating machines themselves, this, limited progress in the field of translation to the development of ideas instead of the actual production of a translating machine or to the practical demonstration of translation .

Modern technology has transposed some of the translator's traditional tools to a new and more flexible medium; more dictionaries (monolingual, bilingual, multilingual) are becoming available in machine readable form. The task of MT can be defined very simply: the computer must be able to obtain as input a text in one language (SL) and produce as output a text in another language (TL), so that the meaning of the TL text is the same as that of SL text.

CONCEPT OF MT:

The term Machine translation (MT) is traditional and standard name for computerized systems responsible for the production of translation from one natural language into another, with or without human assistance. Mechanical and automatic translations are seldom used in English; their equivalents in other languages are still common.

The term does not include computer-based translation tools which support translators by providing access to dictionaries and remote terminology databases, facilitating the transmission and reception of machine-readable texts or interacting with word processing, text editing or printing equipment. It however, includes systems in which translators or other users assist computers in the production of translations, including different combinations of text preparation, on line interactions and subsequent revisions of output. The boundaries between Machine-Aided Human Translation (MAHT) and Human-Aided Machine Translation (HAMT) are often uncertain and the term Computer-Aided (or Computer-Assisted) Translation (CAT) can sometime cover both. The central of MT is the automation of the full translation process.

The translation quality of MT systems could be improved by imposing certain restrictions on the input. The system could be designed, to deal with texts limited to the sub language of a particular subject field such as chemistry.

Alternatively, input texts may be written in a controlled language, which reduces potential ambiguities and restricts the complexity of sentences structures.

Systems are designed either for one pair of language (bilingual systems) or more than two languages (multilingual systems).

AIMS OF MACHINE TRANSLATION (MT):

There have been many different reasons for attempting it. The main reason is a severely practical one: Scientists, technologists, engineering's, economists, administrators, industrialists and businessmen have to read documents and communicate in languages they do not know; and there are not enough translators to cope with the ever increasing volume of material which has to be translated. Machine translation would ease the pressure.

Secondly, many researchers have been motivated by idealism: the promotion of international co-operation and peace; the removal of language barriers; the translation of technical, agricultural and medical material for the world poor. Thirdly, by contrast, some sponsors of machine translation have seen its importance in military and intelligence contexts, to help them to find out what the enemy knows. Fourthly, there are pure linguistic reasons: to study the basic mechanism of language and mind and to find its limitations. Finally, there are simple commercial and economic motives; to sell a successful product.

HISTORY OF MACHINE TRANSLATION (MT):

The interest in machine translation started in the 1950s in the aftermath of the Second World War and was influenced to a great extent by the cold war. During its first decade in the 1950s, interest and support were fuelled by visions of high-speed, high-quality translation of texts especially those of interest to the military and intelligence bodies which funded them quite heavily.

During the second decades (1960s) disillusionment crept in as the number of linguistic mistakes increased. The climax came with the delivery of the National Academy of Science report (ALPAC) in 1966 condemning the field and its workers alike.

As a result machine translation projects were cancelled in the US and elsewhere. In addition, Bar Hillel published his article of disapproval over the project. He concentrated on 'real world knowledge', 'pragmatic knowledge' which machine translation lacked and gave examples "Little John was looking for his toy box.

Finally, he found it. The box was in the pen. John was happy". Here, what is meant by a pen is the playing pen and not the writing pen. Machine translation could not disambiguate the difference in meaning between the two words since it is concerned only with what is in front or it. He gave another example, "The soldiers fired on the women and they fell dead".

MT could not know whether the pronoun 'they' referred to the women or soldiers. By 1973, only three government funded projects were left in the US and by 1975 there were none. Paradoxically, MT systems were still being used by various government agencies in the US and abroad because there was no alternative means of gathering information from foreign (Russian) sources, so quickly. In addition, private companies were developing and selling MT systems based on the mid-60 technology. In the 1980s (the 4th decade), there was a resurgence of interest through the world, plus the growing number of MT and MAT systems in use by governments, business and industry.Industrial firms began to fund (MAT) projects on their own.

By the end of the 1980s machine was firmly established again as an enterprise worthy- of support. Major projects include Eurotra located in several locations across Europe, working on all combinations of the nine languages of the EC. Many companies were funding MT research. In Japan, nearly all the major electronics companies had swiftly developed commercial MT systems, and MT research-notably on speech translation-was also supported by the public sector. The Japanese-funded CICC project (Centre of the International Co-operation for Computerisation) involved five south-east Asian countries in a collaborative project. In the United States-encouraged by European initiatives, the sudden realisation of the strength of Japan and by successful efforts nearer home, including the Pan-American Health Organisation based in Washington DC-MT research at last picked up again, though still not at the level of the late 1950s, and a number of projects were started. The MT community by now had its specialist journals and conferences, and - a true sign of maturity - its factions and arguments.

At this time also, there was a second wave of commercial MT systems, including Siemen's METAL and IBM's LMT, which were both notable in having started out as basic research projects and had completed the full development cycle. Of the newer commercial systems, still some were of simple design, with limited performance, but now the vendors were making more realistic promises. Available systems vary in mode of use and hardware platform, with several aimed at the inexpensive, personal computer end of the market. Almost all commercial systems have quite sophisticated user interfaces, slotting comfortably into word-processing environments alongside writing tools such as spell checkers, thesauri, desk-top publishing packages, and so on (Baker, 1998).

Baker, (1998), summarised the history of MT in an informal way in Figure 1, peaks of activity can be seen in the US just before the ALP AC report, in Japan in the late 1980s when ten or fifteen companies were perfecting their commercial systems, and in Europe in 1990, with the end of the Eurotra disaster, the two later peaks merely signal a change in focus of interest (to other related topics) rather than a loss of funding altogether.

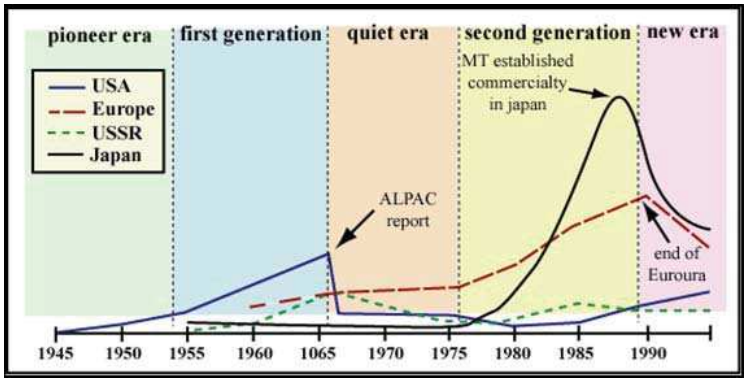

Figure 1. Schematic-graph showing the history of MT; also shown are the five 'Eras of MT History' identified by Hutchins (1992).

The moment MT research activity is on the increase only in the United States and in emerging far-eastern countries such as Korea and Taiwan (Baker, 1998).

Recently, there has been a lot of progress in Artificial Intelligence (AI) and Natural Language Understanding (NLU) has given to Machine Translation.

Current MT systems range from translation aids that facilitate the job of a human translator, instead of wasting time in manual lexicographic searches and in document editing and formatting. Hence, one approach to improving the efficiency of a valuable, experienced human translator is to provide him or her with high powered computational tools for the more mundane, time-consuming tasks. Figure 2 outlines the basic flow of information in a machine aided human translation approach.

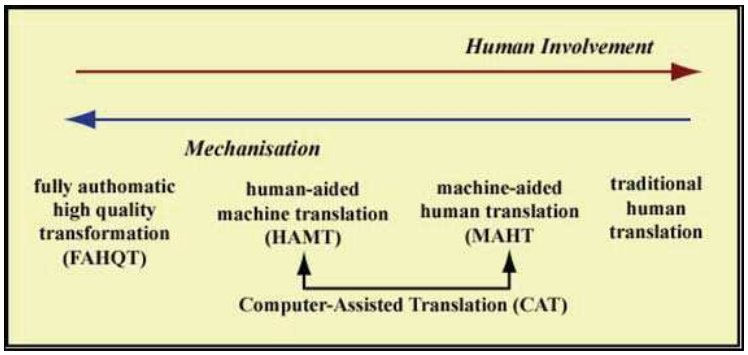

Figure 2. Human and machine translation, After, Hutchins and Somers (1992).

There is a common view, represented diagrammatically in Figure 3, which places human translation and MT at two ends of a spectrum of translation methods with different kinds of human-machine cooperation between them. Computerised systems with no human involved producing translations of a high quality: fully automatic high quality translation (FAHQT).

Human-Aided Machine Translation (HAMT covers the use of MT systems to produce translation with help of translator, before; during or after the computerised processes, whereas

Machine-Aided Human Translation (MAHT) includes the use of computer-based tools as aids for translators.

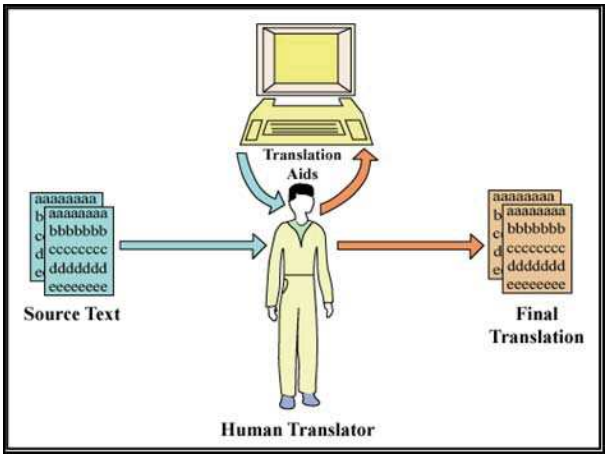

Figure 3. Translation aids, After, Carbonell and Tomita (1987).

HUMAN-AIDED MACHINE TRANSLATION (HAMT):

MAHT, is basically human translation, in other word: it refers to a system where the human is responsible for producing the bulk of translation, but may interact with the system in certain situation; for example, requesting assistance in searching through a local dictionary or accessing a remote dictionary data-bank. The computer does not do the translation; however instead it helps to improve the competency of the translation. It is a known fact that the time taken for translating a source text into a target text (TT) is tool long. Translators spend more time in translation on routines such as dictionary look-up, text formatting, typing-retyping, and checking for terms spelling and grammar and so on. An MAHT system which comes equipped with appropriate tools and subsystems can help reduce the time spent on these chores, as well as maintaining consistency; thus optimising the speed and quality of the translation.

HUMAN-ASSISTED MACHINE TRANSLATION (HAMT):

It refers to a system where in the computer is responsible for producing the bulk of translation, but may interact with human monitor at many stages along the way. Within HAMT there exist two types of systems; first, the system producing translation automatically without human involvement, except before, or after the translation process, and second, the system which translate with human involvement during the translation process. The types known as post-editing processes and interactive translation system and also human may involve outside the process, in pre-editing.

PRE-EDITING:

It basically involves checking source texts (ST) foreseeable problems for the system and trying to eradicate them.Other techniques involve interesting a kind of mark-up in the source texts, indicating explicitly, for example, proper nouns and titles which should be translated, the marking of grammatical categories of homographs, where these might be ambiguous.

This system has the option of marking source texts with special symbols, proper names and indicated by an', sign (e.g. Tom=); (prefix $U), so that the system does not try to analyse them as sentences; a clause without a finite verb can have the prefix SS; and so on.

Pre-editing is optional, which means that the system will look out for these flags, but will still attempt to produce a translation even if they are not there. Pre-editing can improve the overall standard of the translation but it is not indispensable.

Closely related to pre-editing is the use of controlled language in source texts. This approach has been applied with Systran and other systems. The use of controlled language is aimed at adapting source texts to constructions and vocabulary which the system can deal with.

The use of controlled language with MT systems must be distinguished from the sublanguage approach to MT. In latter case, the system itself is designed to deal with the vocabulary and typical constructions of a specific subject area and document type, but there need not be any restrictions on writers or on the texts input to the system. The pre-editing method is illustrated in Figure 4.

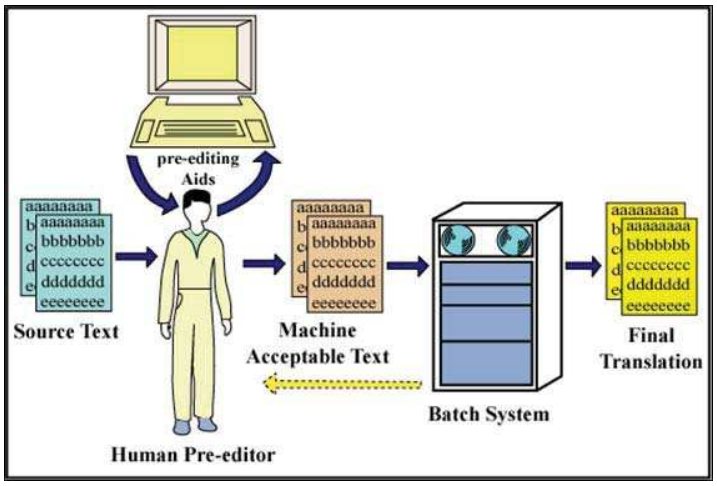

Figure 4. Pre-editing systems, After, Carbonell- and Tomita (1987).

POST-EDITING:

The task of the post-edition is to correct output from the MT system to an agreed standard.

Post-editing was done by hand in the early systems. It consists of tidying up the raw output, correcting mistakes, revising entire, it should be noted that human translations are usually subject to revision, though MT output is quite different from revising human output.

Post-editing alerts the editor to sentences or phrases which may be wrongly translated. Also it provides the option of correcting similar errors automatically throughout the next once the

editor has replaced a mistranslation by a corrected form. As illustrated in Figure 5, systems requiring human post-editing of the translated output. In the following manner:

- The source is converted to computer readable form.
- The text is then sent to a batch-processing MT system, which produces a rough translation several hours or days.
- The original source text and the rough translation are presented to a human translator (the post-editor), who cleans up the translation, fixing any errors or other difficulties.

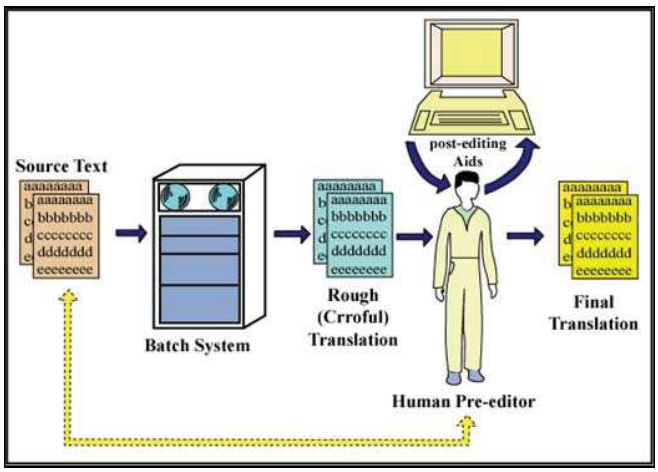

Figure 5. Post-editing systems, After, Carbonell and Tomita (1987).

INTERACTIVE MACHINE TRANSLATION:

The third mode of human assistance occurs in interactive systems. The system consults with the user during the course of performing the translation, clarifying and solving problems of ambiguity in the source text and questions of style in the target text (IT).

In an interactive translation system the translation responsibility is undertaken by the machines, with intervention and help from human (Hutchins, 1986). In the traditional system the machine stops from time to time, that is, in the middle of its translation process, to ask the human user for advice or help. This might be the case where the source text has an unknown word [the user is asked to check its spelling and provide some grammatical information both about the source word (SW) and its translation], a lexical or syntatic ambiguity, or a choice of target translation (IT) either lexical or syntactic.

The difficulty with this kind of interaction is that there are often so many interactions for a single sentence that it might have been quicker for the user to translate from scratch.

Unlike conventional post-editing systems, the interactive approach exhibits the following:

- The user does not have to know the target language (TL).
- The user need not have any special knowledge of linguistics, software, translation.

- The system's final output requires no post-editing.
- Thus, everybody can use the system and generate target language text without assistance of a translator or post-editor. Interactive approach illustrated in Figure 6.

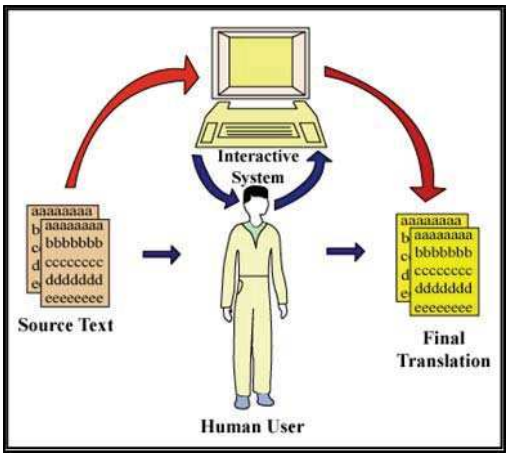

Figure 6. Interactive systems, After, Hutchins and Somers (1992).

TYPES OF MACHINE TRANSLATION:

MT is part of computational linguistics and Artificial Intelligence which explore and use basic language and mind mechanisms by simulation in a computer program, for translation process. There are many models of MT such as Meteo and Eurota:

METEO:

Meteo is one of the major projects of MT. It was introduced in 1977 by Canadian Meteorological Centre, to translate meteorological bulletins from English into French, with a daily rate of 4500 words a day, with an accuracy of up to 90%. Meteo is a second generation MT, however, it is considered a Direct MT type. TAUM group in Montreal developed Meteo system.

The linguistic data of Meteo consist of three bilingual dictionaries for idiom, place, names and general (meteorological) modules for the syntactic analysis of English, the syntactic generation of French, and the morphological generation of French.

EUROTA:

The European community has supported the Eurota project. The project involves about 100 linguists working in the cooperating countries. The framework for translation is proposed as a generalisation of the basic transfer approach. However, the multilingual capability for seven languages requires seven analysis modules 7x6=42 transfer modules, and seven generation modules.

The rationale for this approach is both practical and theoretical Its practical implication is the effort required to develop the 56 modules for seven languages will be significantly less than that which would be needed for developing a separate MT algorithm for each of the 42 different language pairs.

MACHINE TRANSLATION SYSTEM:

Machine translation is a revolutionary idea. It aims at building a machine capable of translating simple and clear texts from one language to another. MT systems have largely developed in three steps. These steps are:-

1. : Direct system
2. : Transfer system
3. : Interlinguas.

DIRECT SYSTEM:

The direct system is an MT approach that is not based on any intermediate systems in the translation process. Direct MT translates directly a text from SL to TL. This strategy is still used partially in indirect MT systems.

Direct MT is considered today as a primitive system. The primitiveness was also partially due to the computers of the 1950's and 60's.

Direct MT system began with a morphological analysis phase. This allows for some identification of word endings and reduction of inflected forms to their uninflected basic forms. The results would be contained in large bilingual dictionaries. However, there is no syntactic structure or semantic relationships. Lexical identification depends on morphological analysis. Figure 9, shows the direct MT system.

Figure 9. direct MT system.

TRANSFER SYSTEM:

Though all systems involve a transfer of one kind or the other, the term transfer system is used to denote a system that interposes bilingual modules between intermediate representations which are language dependent.

The results of the analysis are an abstract representation of the source text, while the input generation is an abstract representation of the target text. The function of the bilingual transfer systems is to render SL (intermediate) representations, as represented in (Figure 10) since these

representations link separate systems (analysis, transfer, generation) they are sometimes referred to as interface representations.

There are no language-independent representations in the transfer system. The source language intermediate is fixed in advance to a specific language, so is the target language.

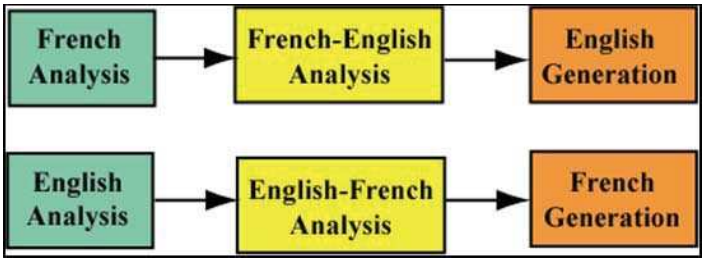

Figure 10. Transfer system.

INTRALINGUA:

The ST is analysed in a representation from which the IT is directly generated. The intermediate representation includes all information necessary for the generation of TT without referring back to the ST. The representation is a reflection from the ST and at the same time acts as the basis for the generation of the TT. It is an abstract representation of the TT as well as a representation of the ST. The system is interlingual in the sense that the representation is neutral between two or more languages as seen in Figure 11.

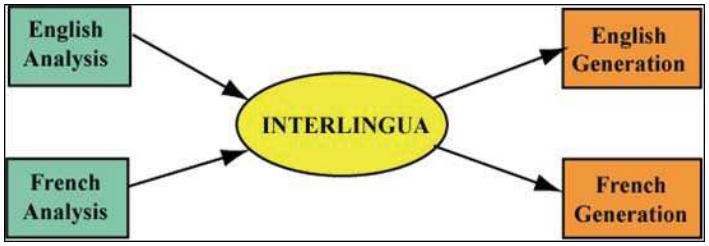

Figure 11. Interlingua systems with two languages pairs.

The advantage of this system is that it is always possible to add a new language and that entails the creation of only two new systems: an analysis grammar and a generation grammar. By adding a German analysis grammar for example, the number of translation directions is increased by four (Figure 12). The building up of Interlingua system can simply continue by adding other languages to it.

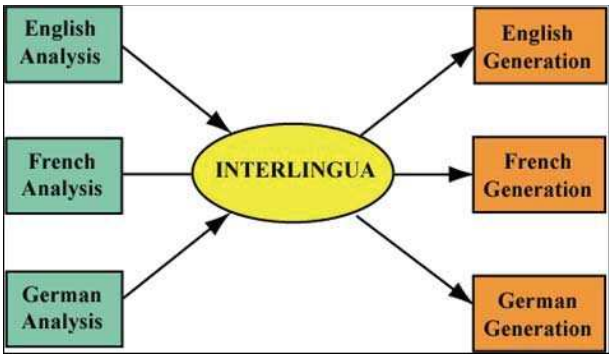

Figure 12. Interlingua systems with three language pairs.

Some Arab countries have taken translation as a means to increase the availability of reading materials in their language (Arabic). But the amount of information and knowledge to be transferred is beyond human capability knowledge are meant for educational purpose. There is a shortage of textbooks for medical students in the Arab countries, even though translation has been adapted as complementary to writing original texts in Arabic translation is found incapable of meeting the needs quickly, Thus the need for an alternative method of producing translation of books, by means of MT. systems.

PROBLEMS OF AMBIGUITY IN MT:

Since lexical or structural ambiguity is part of natural language (NL). In most cases, humans are able to solve lexical and structural ambiguities by using their linguistic abilities as well as world knowledge. However, the problem is that MT does not have such world and linguistic knowledge. To give an illustration of the ambiguity problem:

He went to the bank to deposit his money. A machine translation does not have world knowledge or context knowledge to establish a link: between riverbank and financial bank. The list of such ambiguities and failures is long, but one should look at positive side of MT. Machine translation provides an output that can best be treated as a pre-edited output.

Human translators can then take this output and post-edited... In this way the load of translation work a translator has to do can be reduced. In addition to this, boring and repetitive translation works such as the translation of manuals are best left to MT to translate while human translators dedicate most of their energy to more important and creative tasks. These are some the problems that researchers and translators alike have to face and try to resolve.

STRUCTURAL AMBIGUITY:

Structural ambiguity (SA) involves problems associated with the syntactic structures and the representations of sentences. The following areas will be dealt with:

REAL STRUCTURAL AMBIGUITY:

Real structural ambiguities are said to illuminate alternative syntactic interpretation revealed by formal analysis.

Consider the following example:

Flying planes can be dangerous.

The above sentence it is possible for human readers to find more than a single interpretation, as shown by the following paraphrases:

(la) It can be dangerous to fly planes.
(lb) Planes which are flying can be dangerous.

Flying here can be regarded as gerund governing a complement noun for the first interpretation or an adjective modifying a noun for the second reading.

The ambiguity illustrated in this sentence can be applied to both English and Arabic. In order to get the right meaning of this example when you translate it into Arabic, you should grasp the appropriate meaning and translate it accordingly. Thus it can be translated as the following:

(la) يمكن أن تكون قيادة الطائرات خظرة
(lb) قد تكون من الخطر أن تطير الطائرات

The man saw the girl who possessed the telescope.

There are countless examples of the different ways a given sequence of words may be interpreted that correspond to different structure relations between the parts of the sentence. An account of syntax of the two different representations in the Figure 7 will not satisfactory unless it produces an account of how the different semantic interpretations correlate with the different syntactic relationships that their parts can have to one another.

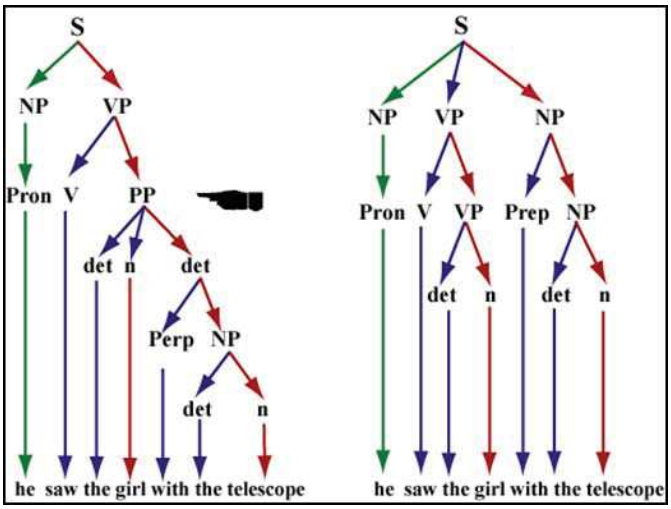

Figure 7. Diagram showing the two different parse trees for: the man saw the girl with the telescope, After, Hutchins and Somers (1992).

The two above diagrams is intended to demonstrate the two different phrase-makers for "the man saw the girl with a telescope". These two trees represent the two different analyses in the sense of recording two very different parse histories. The first can be seen to indicate that the presentational parse (PP) is part of the Noun phrase (NP) that is to say that in this example; it is the girl who possesses the telescope. The remaining option is where the PP is at the same level as the telescope, it is thus the man that has the telescope.

These two different analyses may lead to very different translations, and thus cause problems for man and machine alike.

Similarly, this sentence will cause problem if it is to be translated into Arabic. So to avoid the ambiguity and make the meaning clear the reader has the opportunity to choose one of the possibilities according to the context. He may translate of either as:

(2a) شاهد الرجل الفتاة التي تحمل المنظار

or

(2b) شاهد الرجل الفتاة بالمنظار

ATTACHMENT AMBIGUITY:

Attachment ambiguity is a very predominant feature in MT and arises when there is more than one node to which a particular syntactic constituent may be legally attached.

We have already seen examples of this attachment ambiguity in the following examples:

1: The man saw the girl with the telescope.
2: They complained to the guide that they could not hear.

RELATIVE CLAUSES:

In natural language, a relative clause can be considered as a sentence that contains "gap" which corresponds to head noun- this gives rise to ambiguities of the nature in the example, which also contains different subcategorisation frames for complain and hear:-

1: They complained to the guide that they could not hear.

The word that in the example can be either a relative pronoun (equivalent to whom) such as:-

1. *: They complained to the guide whom they could not hear*.

Or a complementiser as the following:

2. : *They complained to the guide that they could not hear him*.

The word "that" is example of ambiguity as it may be assigned two reading, as a subordinating conjunction.

PREPOSITIONAL-PHRASE ATTACHMENT AMBIGUITY:

Prepositional-phrase attachment ambiguity is the most common example of this phenomenon (Figure 8): phone the man with the limp.

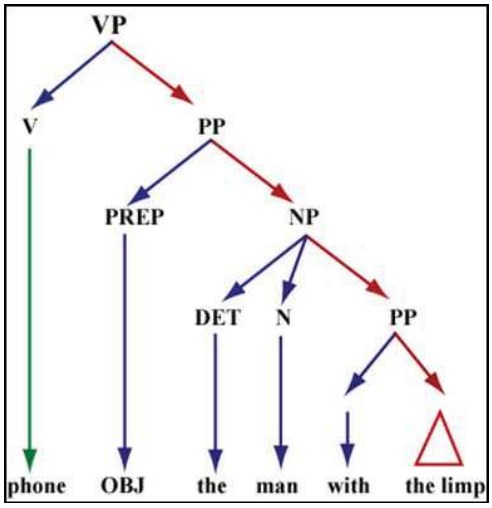

Figure 8. Tree diagram: phone the man with the limp, After Hirst, 1987.

The following sentence is another example of prepositional-phrase attachment ambiguity.

The prepositional-phrase with the telescope may modify either the preceding noun "the girl or the main verb of the sentence "saw". More examples show types of attachment ambiguity.

(1a) Have you read the story about the air crash in the paper?
(1b) Have you read the story about the air crash in the jungle?

The ambiguity lies in attachment of the in-phrase to either the first preceding noun story or the second air crash.

GROSS STRUCTURAL AMBIGUITY:

It arises from the different ways of combining, constituents, for example:

He saw the woman on the hill with the telescope

In this sentence, does "with the telescope" explain how the woman was seen, identify which man, or specify which hill? Consider the following example:

Salsabeel saw the man with the broken glasses

It can be translated into Arabic as:

شاهدت سلسبيل الرجل بالنظارات المكسورة

It may interpreted either:

a. The man who has broken glasses was seen by Salsabeel

or: شاهدت سلسبيل الرجل الذي يضع النظارات المكسورة

b) Salsabeel used broken glasses to see the man.

This is a more difficult process for a machine than for a human, for whom the context of the utterance causes minimal problems. The capacity that they for quickly and easily integrating a sentence into the context of its utterance and extracting from this the information needed to resolve the ambiguity is an almost unconscious procedure for humans but for a machine it is a very difficult process, causing even the simplest of sentences to become massively ambiguous.

CONCLUSION:

This paper gave an overview in machine translation (MT). Most translators produce their translation by means of conventional typewriters, and the output is presented on paper. Since the translation tools used are limited, the output of translation activities is considerably small. Due to these factors, the demand for an alternative means of translation has increased. An alternative means that could help expedite translation, which has been tested and proven successful to a certain extent, is machine translation (MT) systems. Systems such as SYSTRAN, ARIANE, METEO, and a few others are examples of systems that have been developed and utilised quite successfully, despite their limitations, such as knowledge representation were given.

ACKNOWLEDGMENT:

The author wish to acknowledge Department of English, Faculty of Arts, University of Tripoli for utilizing the various facilities in preparation and presentation of this paper.

REFERENCES:

1. Baker, M. (ed.) 1998: Routledge encyclopedia of translation studies, London and New York.
2. Carbonell, G. J and Tomita, M. 1987: Knowledge-based machine translations, the CMU approach, In: Sergei Nirenburg (ed.), Machine translation: Theoretical and methodological issues, Cambridge University Press, London, p. 68-89.
3. Hirst, G. (1987) Semantic interpretation and the resolution of ambiguity, Cambridge, Cambridge University Press.
4. Hutchins, W., 1986: Machine Translation: past, present, future, Ellis Horwood, Chichester, 1986.
5. Hutchins, W., and Somers, H. 1992: An introduction to machine translations, Cambridge University Press, London.