

Lexical and Structural Ambiguity in Machine Translation: An Analytical Study

Yaseen Alzebaree

College of Languages, Nawroz University, Duhok, Iraq

Corresponding Author: Yaseen Alzebaree, E-mail: yaseen.alzebaree@nawroz.edu.krd

ARTICLE INFO

Received: July 20, 2020

Accepted: August 20, 2020

Volume: 1

Issue: 1

DOI:

KEYWORDS

Machine Translation, Lexicon,
Arabic, Challenges, Ambiguity

ABSTRACT

The aim of this study is to investigate the difficulties facing Machine Translation (Google) particularly those related to lexis and structure. The researcher has chosen randomly two English and two Arabic texts about various sorts of translation: Media, Scientific, General and Economic. They were taken from several sources (websites, magazine.) to be translated automatically (Google) and humanly from Arabic to English and vice versa. Then they were analyzed to see the challenges that face Machine Translation (Google). The results of the study indicate that MT is problematic and has many challenges concerning lexis such as (Deletions, non-vocalizations, multiple meanings, collocations, additions and acronyms) and syntax like: word order, verb-subject agreement, passive voice etc. On the basis of the results of the study, the researcher recommended that further work needs to be done to create a system that comprises syntax, morphology and semantics of all languages.

1. Introduction

Machine Translation has come to fill the void existing due to the small number of translators and the impossibility of getting them everywhere. It saves both time and money. A large quantity of articles and documents are easily translated in a short time with a low amount of money but it encounters many problems. Technical revolution has led to the development of knowledge fields in general. Human beings dependence on the machine is increasing day by day. Using machine in work has become one of the advanced phenomena, and its absence from work is considered as retardation and poor production phenomenon. Using machines in factories and other places of works has many advantages including speed in achievement, abundance and accuracy. Computer is one of these machines and it is the top product of mankind up to now which is available nearly in all work places, but the area that forms a challenge for this device is the area of language. In recent years, Machine Translation (MT) has been given special attention and concentration by researchers and scientists due to many factors and reasons, the most important of which is the increasing need of communication between people living in different parts of the world and speaking different languages. The need for machine translation has been felt since the advent of computers, but we can say that the early system attempts of machine translation have proved to be dissatisfactory as, Attia (2002) says that machine translation was based on a primitive idea of processing the source text through an electronic dictionary that included words of the source language and their equivalents in the target language, with no further manipulation either of the input or the output.

The process of translation through the use of machine or computer between human languages is not an easy job or task since human language has a highly complicated system. In spite of the great progresses that the computational linguistics has witnessed, MT is not perfect and is still far from being satisfactorily accomplished and has a long way to go.

There has been a need of translation of information among different languages for thousands of years. Translation requires not only advanced skills in the source and target languages but also knowledge of culture of both languages, especially in the literary translation, as Şenkal (2000) claims that literary translation refers to poetry, drama and other literary works from one language to another, the ability to choose the correct translation of an element given a variety of factors is vital. By contrast,

most of the translations in the world do not contain a high level of literary and cultural knowledge. The majority of professional translators are working on translations of scientific and technical documents, commercial and industrial transactions, legal documentations, instruction manuals, technical and medical text books, industrial patents, news reports, etc.

Furthermore, the demand for language translation has greatly increased in recent years due to increasing cross-regional communication and the need for information exchange. Most materials need to be translated, including scientific and technical documentations, instruction manuals, legal documents, textbooks, publicity leaflets, newspaper reports. (Karamat, 2006). It is becoming difficult and impossible for professional translators to meet the increasing demands of translation. Therefore, we see in such a situation that the assistance of computers can be used as a substitute. Machine translation (MT) is an important technology for converting information from one language into another with the help of a computer. However, the large number of languages prevalent in the world makes translation a huge task. (Kumar, 1994)

The researcher himself, being a student of translation, has used machine translation (especially Google translation) too much and has seen that it has, in spite of its advantages, many faults and disadvantages, among these faults are syntactic and semantic problems, etc. There are many difficulties and problems that prevent MT from being able to translate a text perfectly. Among these difficulties and problems are lexical and structural ambiguity which means that a word or a text can have two or more meanings. Ambiguity comes in two forms; it is either lexical or structural. Most output is unsatisfactory as compared with human translation and needs to be edited.

There is no doubt that Machine Translation has played an important role which cannot be ignored that is to convey sufficient information from SL to TL. One can gain necessary information from it and the existence of MT helps in bringing down the communication barriers. It is obvious nowadays that more translators are required than before. MT is important for a variety of reasons. Human translation is expensive, takes time and is usually unavailable when it is needed for communicating quickly and cheaply with people whom we do not share a common language. Due to the advantages mentioned above about MT, we see that it is necessary to point out to its disadvantages.

2. Literature Review

The importance of translation has increased recently and its role has expanded in various international organizations that require translation of government documents from and to the other languages as noted in the European Common Market, which spends more than half of its budget on the translation. Furthermore, scientists need a quick translation of the reports and the results of research published by their peers in other countries. The intelligence agencies of several countries also translate the vast amounts of documents and information. These organizations are facing difficulty to get qualified translators who can translate this vast amount of materials. One may ask why machine translation exists. The first and most important reason is that human translation can't deal with daily massive production of data which human translation cannot cope with it especially after the outbreak of the information revolution and the diversity of languages, which increased their number over 4000 languages in all parts of the world today hence we see that MT is a necessity.

Advances in technology and in telecommunication, especially in computer science, have made possible the access to an enormous amount of data, including texts, almost instantaneously anywhere in the world. And information systems are making it easier to find the material we require amongst a huge store of data. Converting printed text into machine-readable data, performing complex searches in databases which might be situated on the other side of the world, and obtaining instant responses are among the computer advances that we can now have the use of.

To many people, today's information technology revolution is as extraordinary, and has had such great impact, as the invention of printing had in other periods of history. Actually, printing, photocopying and other technological developments of the kind have helped writers in their works and their distribution; but the Information Technology (henceforth IT) revolution not only

helps the writer in a mechanical sense to produce a neat and attractive final version, but makes it easier to create texts and provide search and analysis tools.

Translating between languages is as old as mankind, and it is extremely widespread, taking into account that there are more than two thousand different languages in use today. Linguistics information technologists have dreamt the automation of translation for many years. Though it is still a defective one now, but this dream has come true and became a reality. In fact, researchers have been working on machine translation of languages for 50 years now; and the successes obtained, as well as the good number of commercial systems in operation today- whether as institutional work stations or as products available on the market for the individual PC user- show that Machine Translation is more than a mere imaginative possibility; though high-quality fully automatic machine translation remains an elusive goal. (Guerra, 2000)

In the present 'information age', computers are being used for processing, storage and retrieval of information. At the same time, the need for sharing of this information among all countries of the world has also been felt. Faster communication techniques are being used for easy portability of information. But the existence of information in different languages still remains one of the barriers in its use world-wide.

Machine translation (MT) systems have, in some form or another, been in development for quite a few decades, but to this day there are no systems which are able to produce flawless translations. "Flawless" is certainly a subjective concept, but the truth is that machine translation still has a very long way to go. Some people may wonder why this is so. If a computer can perform calculations impossible for a human in a matter of seconds, why would it not be able to translate a text, which most humans could translate without much difficulty (Hoppe, June 2005).

The concept of translation

Translation is a profession besides being an art. Newmark defines translation as "the attempt to produce approximate equivalence or respectable synonymy between two chunks of different languages on various levels, of which the two main ones are thought and linguistic form." He adds that translation is partly an exercise in the art of writing as well as a field of comparative applied linguistics. Newmark further says that the word, not the sentence, is the essential element of translation. (Hamdalla, 1998: 24). (qtd. in Newmark, 1973)

Methods of translation

For the purpose of understanding the *concept* of machine translation, it is desirable to examine various methods of translation. They are as follow:

Human Translation

Human translation is as old as mankind and the most widely used method. It is also the most reliable method. The other three are considered to be synonymous and are labeled as MT. Some draw a distinction between MAT (also called computer assisted translation) and MT. In MAT, the role of the machine is to assist the human translator, as it is obvious. On the other hand, in the human-aided machine translation, the machine, besides assisting, does the translation as well. The MT is the ideal case that falls in the domain of artificial intelligence. (Lehrberger & Bourbeau, 1984: 6-8)

Machine-Assisted Human Translation (MAHT)

MAHT is basically human translation with only limited assistance from the machine. At the lower end of the scale of what might be called "computerized translation" the machine may consist simply of a word processor with provision for looking up translation equivalents of source language words. This may be faster than writing out the translation by hand (or typing it with an ordinary typewriter) and thumbing through a dictionary for unfamiliar terms, but it doesn't remove from the translator the burden of actually performing the translation. Following are some features that may be included in an MAHT system. (ibid)

Human-Assisted Machine Translation (HAMT)

In the case of HAMT, the human translator supplies limited information to "fill out" the machine translation. After being supplied with the necessary data by the translator, the machine completes the translation, producing a raw output suitable for human revision. (ibid)

Fully Automatic (Automated) Machine Translation (FAMT).

In FAMT, there is no human intervention between the input of the original text and the final raw machine output of the translated text. Of course, revision of the raw output may be required, just as it is for the output of a human translator; "fully automatic" does not imply that human post-revision is eliminated. However, there is at least one FAMT system in which the machine itself decides which of the sentences submitted to it are to be revised, all others being translated and considered suitable as finished text ready for use.(ibid)

What Is Machine Translation?

Daxbock (2010: 2) defines MT as a subfield of computational linguistics, which investigates the use of computers to translate text or speech from one natural language to another. The ultimate goal is to program computers to understand text or speech in source language (SL) like a human does, to create output in target language (TL) that sounds as if it were from a human.

A Brief History of Machine Translation

The idea of using computers to translate or help translate human languages is almost as old as the computer itself (Trujillo A. 1999). There are certain eras that MT has gone through:

Pre-computer: Some of the ideas that have influenced MT were already current or at least existent in the pre-computer era. Since at least the 17th century, scholars and philosophers have proposed the use of language-neutral representations of meaning in order to overcome linguistic barriers.(Mohamed, 2000:18-20)

Initial efforts: Early proposal for the use numerical techniques in MT can be traced at least to 1947, when computers had just been successfully employed in deciphering encryption methods during the Second World War. A memo from Warren Weaver proposed specific strategies for using computers to translate natural languages. This memo initiated MT research in the USA and in the rest of the world, with the first public demonstration of a Russian-English prototype MT system in 1954. This event led to similar work in the then USSR and other places around the world. (ibid)

In the 1950s, it was realized how powerful and efficient computers were in performing complex mathematical calculations. This led researchers to believe that computers could also easily be used to translate natural languages. They set out to create such a system and tried to formulate precise rules about lexicon and grammar which the computer would understand. This, however, proved to be much more difficult than they had originally anticipated. (ibid)

The development continued, however, and in 1954, MT system was presented at the head office of IBM. The project generated a great deal of public interest but was in fact a very simple system, consisting of 250 words and able to translate 49 selected sentences from Russian into English. This led to increased financial support for the field of MT (ibid). During the Cold War, a system had been developed in the US, which was used for translations of articles in Russian scientific journals. Although the translations generated by this system were still very rough, it was possible to understand the essence of the articles. If they were deemed interesting from a security perspective, they were forwarded to a translator. (ibid)

The Automatic Language Processing Advisory Committee (ALPAC) Report (1966): The initial over-optimism in MT came to an end in the USA when the ALPAC report, commissioned by government sponsors of MT, suggested that MT was not cost-effective.

The 1970s and operational MT: Continued effort in MT yielded operational systems in the early 1970s. Systran began Russian-English translations for US Air Force in 1970, while Meteo began translating weather reports in 1976. Also in 1976, the commission of the European Union (then communities) installed an English-French version of Systran.(ibid)

Rebirth in the early 1980s: The late 1970s and early 1980s saw an increase in interest in MT. The Eurotra project from the European community began in the 1982, influenced by work done at Grenoble and Saabruken since the 1960s and 1970s. Similarly, in Japan the Mu project started in 1982, and knowledge-based MT started in 1983 in the USA. Some commercial systems also began to appear.

Late 1980s and early 1990s: A number of companies, especially large Japanese electronics manufactures, began to market MT software for workstations.(ibid)

Late 1990s and Machine Assisted Translation MAT: At the end of the decade we saw powerful translation engines on personal computers, translation on the Internet, widespread use of translation memory and translator's workbenches, multimedia and software localization, as well as an increased interest in Example-based MT (ibid)

Approaches to MT

MT Systems are designed either for two particular languages (bilingual systems) or for more than a single pair of languages (multilingual systems). Bilingual systems may be designed to operate either in only one direction (unidirectional), e.g. from Japanese into English, or in both directions (bidirectional). Multilingual systems are usually intended to be bidirectional; most bilingual systems are unidirectional (Hutchins, 1995:1). Three main linguistic approaches to Machine Translation are surveyed, namely direct, transfer, and Interlingua. The objective of this section is to give an overview of what these linguistic methods are, how they work, and what their advantages and disadvantages are. (Hui, 1998, p.3). There is another approach which is spoken-language MT.

Direct

In overall system design, there have been three basic types. 'Direct translation' approach is the first (and historically oldest) type. The MT system is designed in all details specifically for one particular pair of languages, e.g., Russian as the language of the original texts, the source language, and English as the language of the translated texts, the target language. (Hutchins, 1995: 2). The direct translation approach to MT is the most basic and least sophisticated one. Dictionary entries are used to translate word by word. (Daxbock, 2010: 3). Processing in MT systems that follow the direct strategy consists of three stages:

1. **Morphological analysis of the source language input text.**

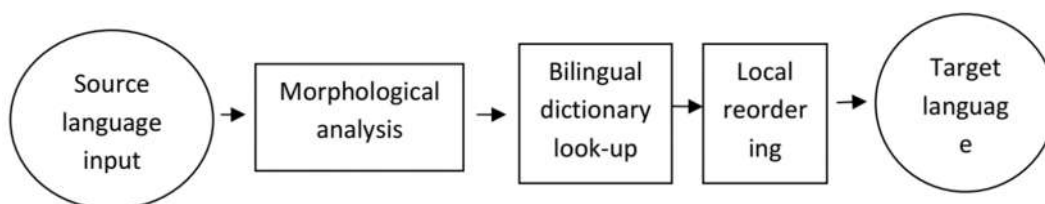
In this stage, the system identified word ending and reduces inflected forms to their uninflected base forms.

2. **Bilingual dictionary lookup.**

Depending on a huge bilingual dictionary, the system decides the correct replacement for source words with equivalent words in the target language.

3. **Local reordering of the target language.**

After the replacement is done, the system makes adjustment to the output text by applying rules for putting words in their right order. The following figure summarizes this process.



This approach has advantages and disadvantages. It is easy to implement for bilingual translation because the basic idea is to look up each word in the bilingual dictionary and select a corresponding translation. However, the accuracy is low, because of ambiguity in word meanings and the lack of syntactic knowledge in the system. Furthermore, it is impractical to build a multilingual system based on the direct method (Hui, 1998: 7). Despite the clear disadvantages of the direct method, it is still applied to some extent in unidirectional bilingual systems. These systems “take advantage of similarities of structure and vocabulary between source and target languages in order to translate as much as possible according to the direct approach (Attia, 2002: 13-14).

Transfer

The second basic strategy is the less ambitious transfer approach. Rather than operating in two stages through a single interlingual representation, there are three stages involving underlying (abstract) representations for both SL and TL texts. The first stage converts SL texts into abstract SL-oriented representations; the second stage converts these into equivalent TL-oriented representations; and the third generates the final TL texts. (Hutchins, 1995:2). The inaccuracy of the direct translation mechanism led to the notion of an intermediate representation of meaning. This idea led to a new approach, called transfer. Although semantics is incorporated into this approach, the representation is still language dependent.

From figure 2 below, we see that both SL and TL require their own analysis module as well as their own generation module. This entails that given SL, we need to analyze it, apply language dependent rules, and generate TL. (Hui, 1998: 3-4)

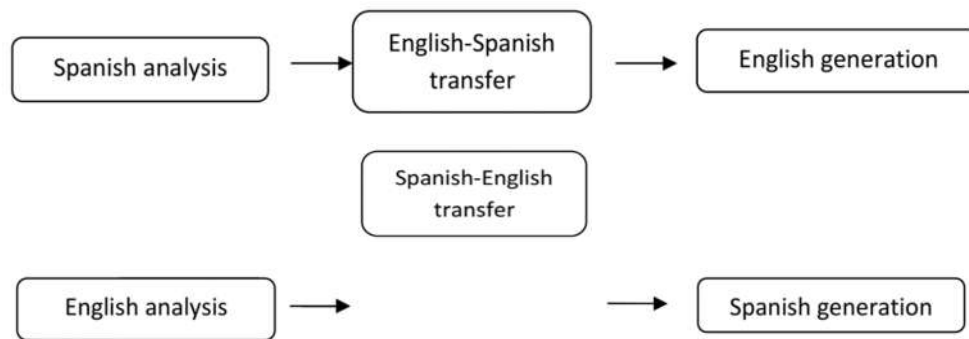


Figure 2: Transfer MT System

The transfer method is a middle course between direct and interlingua MT strategies. The difference between the three strategies can be captured in the following figure:

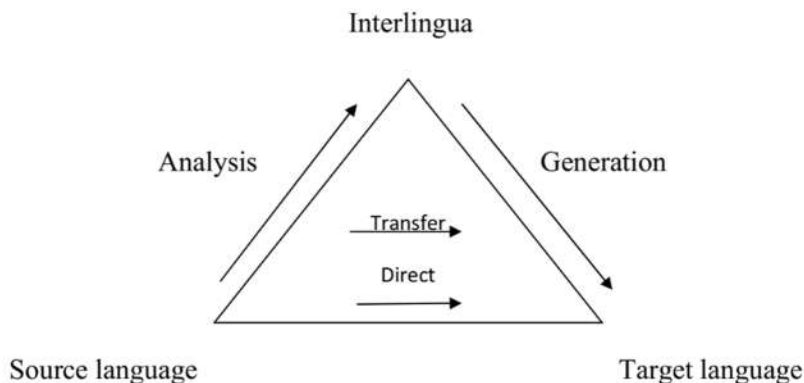


Figure 3: Difference between direct, transfer, and Interlingua MT methods

Interlingua

The third basic design strategy is the Interlingua approach, which assumes that it is possible to convert SL texts into representations common to more than one language. From such interlingual representations texts are generated into other languages. Translation is thus in two stages: from SL to the interlingua (IL) and from the IL to the TL. (Hutchins, 1995:2).

Interlingua is an idealistic approach to solving the machine translation problem because it relies on conceptual representation and linguistic principles. Unlike transfer, the intermediate representation is language independent. This feature makes Interlingua theoretically most attractive because it is the closest technique to arriving at a universal solution. Figure (4) below illustrates the architecture of an Interlingua system. (Hui, 1998:6)

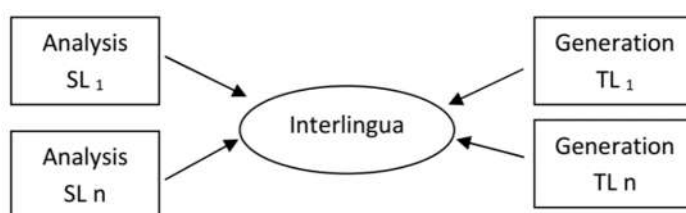


Figure 4: Interlingua MT System

Spoken-language MT

Spoken-language MT is more than a speech-to-text front and backend in a conventional text MT system. Spoken-language is hugely different from written language. The basic problem is to isolate the speech signal from the surrounding noise. Other speech related problems are hesitations, self-repairs during speech, accents, mixed language speech, use of anaphora and ellipsis. The dominant form of spoken language MT used to be the translation of dialogues (Daxbock, 2010:9).

Why does Machine Translation matter?

Human translation cannot cope with the daily massive production of data. It cannot either deal with the technical material and terminology that are used consistently and need to be translated in the same way every day. Consequently, this leads us to envisage machine translation (MT) as probably the only reliable solution. MT is not in itself an independent line of work; it draws from linguistics, computer science, artificial intelligence and translation theory. There is no doubt that Machine Translation has played an important role in transferring information from a language to another and will continue to evolve, given that we are living in a multi-cultural and multi-lingual environment. The existence of MT helps in bringing down the communication barriers.

With the dominance of English over other languages in many fields, it is obvious that more translators are required than before. The amount of translations carried out from English toward other languages is vast and is worth billions of dollars. The amount of translations carried dollars. The growth of the Internet and the computerization of the worldwide economy have changed the

manner business is being conducted, emphasize the call for more efficient and faster techniques of translation, and manage to make use of the huge volume of available data online. (Bessou & Touahria, 2011)

MT is undoubtedly an important topic socially, politically, commercially and scientifically, and intellectually or philosophically.

The *social* or *political* importance of MT arises from the socio-political importance of translation in communities where more than one language is generally spoken. Translation is necessary for communication — for ordinary human interaction, and for gathering the information one needs to play a full part in society. Being allowed to express yourself in your own language, and to receive information that directly affects you in the same medium, seems to be an important, if often violated, right. And it is one that depends on the availability of translation. The problem is that the demand for translation in the modern world far outstrips any possible supply. Part of the problem is that there are too few human translators, and that there is a limit on how far their productivity can be increased without automation. In short, it seems as though automation of translation is a social and political necessity for modern societies which do not wish to impose a common language on their members. (Arnold et al., 2001).

This is a point that is often missed by people who live in communities where one language is dominant, and who speak the dominant language. Speakers of English in places like Britain, and the Northern USA are examples. However, even they rapidly come to appreciate it when they visit an area where English is not dominant (for example, Welsh speaking areas of Britain, parts of the USA where the majority language is Spanish, not to mention most other countries in the world). For countries like Canada and Switzerland, and organizations like the European Community and the UN, for whom multilingualism is both a basic principle and a fact of everyday life, the point is obvious. (ibid)

The *commercial* importance of MT is a result of related factors. First, translation itself is commercially important: faced with a choice between a product with an instruction manual in English, and one whose manual is written in Japanese, most English speakers will buy the former. Secondly, translation is expensive. Translation is a highly skilled job, requiring much more than mere knowledge of a number of languages, and in some countries at least, translators' salaries are comparable to other highly trained professionals. (ibid).

Scientifically, MT is interesting, because it is an obvious application and testing ground for many ideas in Computer Science, Artificial Intelligence, and Linguistics, and some of the most important developments in these fields have begun in MT. (ibid)

Philosophically, MT is interesting, because it represents an attempt to automate an activity that can require the full range of human knowledge — that is, for any piece of human knowledge, it is possible to think of a context where the knowledge is required. For example, getting the correct translation of *negatively charged electrons and protons* into French depends on knowing that protons are positively charged, so the interpretation cannot be something like “negatively charged electrons and negatively charged protons”. In this sense, the extent to which one can automate translation is an indication of the extent to which one can automate ‘thinking’. (ibid)

Vagueness

Though seemingly synonymous in common usage, vagueness and ambiguity are entirely different but very important problems in critical thinking. The difference between ambiguity and vagueness is enumerability: an expression is ambiguous if its interpretations can be enumerated as a finite disjunction of potential meanings; an expression is vague if its meaning cannot be given as a finite disjunction of possibilities. Or to put it in simpler terms:

- A word or phrase is said to be ambiguous if it has at least two specific meanings that make sense in context.
- A word or phrase is said to be vague if its meaning is not clear in context. (Corr, 2005:8)

Machine Translation Problems

Izwaini (n.d., p. 119-135) states that there are many problems in the output of English into Arabic translation mode. These are divided into lexical and grammatical problems:

Problems of Lexis

The main problems related to lexis in this mode are the lexicon is inadequate and not updated, polysemy and wrong TL word, connotation and collocation. There are also two problems of addition and deletion.

Ambiguity

The relationship between ambiguity and language is both intrinsic and interdependent. The fact is without one we cannot have the other. Ambiguity exists in various levels of seriousness and is one of the features of human language that makes it so unique and complex. Humour, one of the more playful and creative capabilities of language, is based on ambiguity. (Corr, 2005, abstract, 8).

‘Ambiguity’ is the term for the existence of at least two separate, incompatible interpretations of a stretch of speech. Two kinds of ambiguity are recognized: if it is due to words, it is called ‘lexical’ ambiguity; if it is due to sentence structure, it is called ‘structural’ ambiguity. (Strazny, 2005:40).

Lexical Ambiguity

In this chapter, we turn our attention to an issue we set aside earlier: the fact that in different contexts the same word form can be understood to have different semantic interpretation. For instance, a book can mean **'a collection of pages that are bound in a rigid cover'** (hand me that book), **'the information contained in a collection of pages that are bound in a rigid cover'** (that book is depressing) or **'reserve'** (We want to book a table for four). (Murphy, 2010). Lexical ambiguity can be either ‘polysemy’ or ‘homonymy’.

Polysemy

The term 'polysemy' is derived from the Greek *poly-* meaning 'many' and *sem* meaning 'sense' or 'meaning'. In its simplest form, then, the term refers to the phenomenon in language where one linguistic form has a number of different, yet related meanings. (Cuyckens & Zawada, 2001, p. x).

Homonymy

Homonyms are unrelated senses of the same phonological word. Some authors distinguish between homographs, senses of the same written word, and homophones, senses of the same spoken word. Here we will generally just use the term homonym. The term homonymy goes back to the Greek words 'homos', meaning 'the same', and 'onoma', meaning name. A simplified translation of homonymy would be 'having the same name'. (Schulze, 2001).

Multiword Units: Idioms and Collocations

Idiomatic expressions pose a particular challenge for the today’s Machine Translation systems, because their translation mostly does not result literally, but logically. (Anastasiou , 2010: abstract).

A collocation is an expression whose meaning cannot be worked out from the meanings of its constituent words. Even if you know the meanings of all the words in the phrase **let the cat out of the bag**, you cannot guess the idiomatic meaning of the whole expression: this you must learn separately. (It means, of course, to reveal something publicly which is supposed to be a secret.) (Trask, 2007:114).

The translation of collocations is difficult for non-native speakers of a language. Differences in the collocational patterning of the source and target languages create potential pitfalls and can pose various problems in translation. The difficulty for the learner is not to recognize a source language collocation but to find an acceptable target language equivalent.

The translation of collocations is an ever-lasting struggle for learners of English to match the proper noun with the proper verb or adjective. The problem is that language is not made up of a large number of words that can be used together in free variation; words have a certain tolerance of compatibility. Like individual words, collocational patterns carry meaning and can be culture-specific; this in addition to their largely arbitrary nature, gives rise to numerous problems in translation. (Boussalia, 2010)

Acronyms

Words made up of initial letters are known as acronyms, of which well-known examples are NATO (for North Atlantic Treaty Organization), RAM (for Random Access Memory), and AIDS (from Acquired Immune Deficiency Syndrome). Intermediate between an acronym and a blend is sonar (from Sound Navigation and Ranging). (McCarthy, 2002:65)

Structural Ambiguity (syntax problems)

Structural ambiguity occurs when a phrase or sentence has more than one underlying structure. The main problems related to grammar and syntax in the English Arabic mode are word order, gender, reference, tense and aspect, prepositions, the definite article, and wrong analysis of grammatical categories. (Izwaini, n.d.:119-135)

3. Methodology

In order to conduct this research, a set of texts about different types of translation are selected to be translated automatically through (Google translation) and humanly (human translation) and then analyzed to see if there are any problems.

3.1 Study Sample

In doing the project, the researcher has collected the required data which consists of two Arabic texts and two English texts about different types of translation: Media, Economic, Scientific, and General Translation. These texts are chosen randomly from different sources: Magazine, Press, and Website, the Associated Press, Almujtamaa Magazine, Medical News Today, and alarabiya.net.

3.1.1 First Sample (Economic Text)

The first sample, a piece of Arabic text translated into English, is about economic translation taken from Alarabiya.net on December 25, 2011 under the title: **الرؤود العالمى ىجبىر "الأمم المتحدة" على خفض ميزانيتها** taken from <http://www.alarabiya.net/articles/2011/12/25/184409.html>.

3.1.2 The Second Sample (General Text)

The second sample, a piece of Arabic text translated into English, is about general translation taken from Almujtamaa Magazine. The text entitled **"اكتتاب وعدوانية وجرانم عانلثة الجنود الأمريكوى العاندون من العراق"**. It is dated on December 24, 2011. Its issue No. is (1982). <http://magmj.com/index.jsp?inc=5&id=8504&pid=2131&version=122>.

3.1.3 The Third Sample (Scientific Text)

The third sample, a piece of English text translated into Arabic, dealt with scientific translation taken from Medical News Today on 02 January 2012 under the title: "**Majority of Smokers do Not Appreciate the Risks**" through: <http://www.medicalnewstoday.com/articles/239800.php>.

3.1.4 The Fourth Sample (Media Text)

The fourth sample, an English text translated into Arabic, is about Media translation taken from the Associated Press under "**Libya marks 1st Independence Day in 42 years**" issued on Saturday, 24 December 2011. http://www.cbsnews.com/8301-202_162-57348202/libya-marks-1st-independence-day-in-42-years/

3.2 Study Tools

The researcher has used two tools: the first one is the "Google" translation and the second one is the "Human" translation.

3.3 Procedures

The researcher used the following procedures during the application of this study:

- The researcher has selected four different texts about different kinds of translation (general, scientific, media and economic) in both English and Arabic languages taken from different sources: Website, Magazine and Press.
- The texts have been translated automatically (Google translation) and humanly (the researcher himself).

3.4 Validity of the Instrument

For validating the translation of texts, the researcher translated them from English into Arabic and vice versa then they were submitted to Dr. Babikir Gorashi specialized in translation from the Faculty of Languages, University of Sudan for Science and Technology. After approving their suitability for the purpose of the study, the researcher took the suggested modifications into consideration and then analyzed the texts.

This study has the following limitations:

1. The researcher has selected four different texts about different kinds of translation (general, scientific, media and economic) in both English and Arabic languages taken from different sources: Website, Magazine and Press.
2. The texts have been translated automatically (Google translation) and humanly (human translator)

4. Results and Discussion

Problems of Lexis

Deletion

One might argue that in the MT context, deletion is a technical problem rather than a linguistic problem, in that it is related to the inability of the system to read and recognize words. Nevertheless, it causes a linguistic problem in the output. (Izwaini, n.d, : 120). There are some cases of deletions by Google, which are peculiar since they are content words what are deleted. Some words are dropped with no apparent reason for such a procedure. For example, "وعدها 193 بلدا" is translated into "*and the 193 countries*" while it should be "*it counted 193 countries*".

"العام المالي الماضي" is translated into "for the fiscal year" while it should be "for the last fiscal year"

"وقال الامين العام للمنظمة الدولية بان كي مون" is translated into "The secretary General of the International organization Ban Ki-moon" while it should have been translated as "The secretary General of the International organization Ban Ki-moon said"

"الدول الاعضاء" is translated into "Member States" while it should be "The member states"The is omitted here.

"Long-term smokers die" is translated into "على المدى الطويل يموت". It should be "المدخنين لمدة طويلة يتعرضون للوفاة". *Smokers* have been deleted in MT. All the above deletions can be shown in the following table.

Table No. (1) Deletion

Original Text	MT	Human Translation
وافقت الدول الأعضاء بالأمم المتحدة، (Economic Translation) وعدها 193 بلدا.	Member States approved the United Nations, and the 193 countries	The member States in the United Nations, which counted 193 countries, agreed
للعام المالي الماضي (Economic Translation)	for the fiscal year	for the last fiscal year
وقال الامين العام للمنظمة الدولية بان كي مون. (Economic Translation)	The secretary General of the International Ban Ki-moon	The secretary General of the International Ban Ki-moon said
Long-term smokers die. (Scientific Translation)	على المدى الطويل يموت	المدخنين لمدة طويلة يتعرضون للوفاة

Addition

There are some cases of additions by Google. Some words are added with no clear reason for such a procedure as it is shown in the following examples:

"ووصف المفاوض الامريكي جوزيف تورسيلا الاتفاق المتعلق بالميزانية بالتاريخي" is translated into "**He** described the U.S. negotiator Joseph Tursala historic, agreement on the budget". It should be "The U.S. negotiator, Joseph Tursala, described the agreement on the budget as historic". It appears here that word order creates problems.

"واعترف بان كي مون" is translated into "**He** admitted that Ban Ki-Moon". It should be "Ban Ki-Moon admitted".

"More than half of smokers underestimate" is translated into "اكثر من نصف المدخنين **عن التبخين** يقلل". It should be "اكثر من نصف المدخنين يستخفون". The following table shows additions.

Table No. (2) Addition

Original sentence	MT	Human Translation
ووصف المفاوض الامريكي جوزيف تورسيلا (Economic Translation)	He described the U.S. negotiator Joseph Tursala	The U.S. negotiator, Joseph Tursala, described
واعترف بان كي مون (Economic Translation)	He admitted that Ban Ki-moon	Ban Ki-Moon admitted
more than half of smokers underestimate (Scientific Translation)	اكثر من نصف المدخنين عن التبخين يقلل	اكثر من نصف المدخنين يستخفون

Non-vocalization

Non-vocalization is a major cause for mistranslations. Having no diacritics renders many words homographs, and thus it is difficult to determine which meaning to choose. For example, "أقدم" can mean "oldest" or "did something" but the context in which "أقدم" came in means "did something" while Google translated it into "oldest". Here Google translation faced a challenge of non-vocalization and failed to transfer the meaning correctly as it is shown in the table below.

Table No. (3) Non-vocalization

Original Text	MT	Human Translation
حتى أقدم على إزهاق روح زوجته. (General Translation)	Even the oldest on the spirit of the loss of his wife	he killed his wife

Multiple Meaning

Another major problem is the word that has more than one meaning. Many English words, for example, can have two or more meanings in Arabic language and vice versa. As a result, problems arise as MT systems can't decide which one to choose.

Homographs

Another challenge encountered by MT is homographs. They are two (or more) 'words' with quite different unrelated meanings which have the same spelling. The word "نالت" has a negative meaning here in the following sentence: "نالت اللعنة المجتمع الأمريكي" but Google has translated it differently into "but won the curse of American society". It should be "but it inflicted curse on the American society"

Another example is the word "marks" in "Libya marks 1st Independence Day in 42 years". It is translated into "علامات" which is different and wrong than it should be in this context. It should be "تحتفل", which doesn't exist in the lexicon of Google at all. Here the problem of inadequate lexicon arises which means some basic words are not included in the dictionary.

The word "mark" was translated in another context into "اتسمت", while here it means "لحيون" in this context. And the word "article" was translated into "المادة" rather than it should be in the context it comes which means something else "مقالة". The word "launched" in "which has launched a new Smoke free campaign" is translated into "شنت", while it should have the word "بدأ" as Arabic equivalence.

The word "ناله" in the following context "وأول من يناله ذلك الجانب هم عائلاتهم" carries negative meaning but MT has translated it positively. It was translated into "And the first to be bestowed on that side are their families", while it should be "The first to be affected by that behavior are their families". The following table shows homographs:

Table No. (4) Homographs

Original sentence	MT	Human Translation
نالت اللعنة المجتمع الأمريكي. (General Translation)	but won the curse of American society	it inflicted curse on the American society
Libya marks 1st Independence Day in 42 years. (Media Translation)	ليبيا علامات 1 يوم الاستقلال في 42 سنوات	للمرة الأولى تحتفل ليبيا بيوم استقلالها الأول منذ 42 سنة
Only the 1969 date of his coup was marked . (Media Translation)	اتسمت فقط لعام 1969 تاريخ انقلابه	لحيون ذكرى 1969 تاريخ انقلابه فقط

article date (Scientific Translation)	تاريخ المقالة	تاريخ المادة
which has launched a new Smoke free campaign. (Scientific Translation)	اطلقت حملة جديدة للتخلص من التدخين	التي شنت حملة جديدة الخالية من الدخان
وأول من يأثله ذلك الجانب هم عائلاتهم. (General Translation)	The first to be affected on that behavior are their families	And the first to be bestowed on that side are their families.

Collocations

Translation of collocations is difficult for nonnative speakers as it is difficult to find TL equivalents. MT (Google) also faces problems in translating collocations. Many collocation translations are idiosyncratic in the sense that they are unpredictable by syntactic or semantic features. The following table shows homographs:

Table No. (5) Collocations

Original Collocation	MT	Human translation
long-term smokers (Scientific Translation)	طويلة الأجل المدخنين	المدخنين لأمد طويل
long-term smokers die (Scientific Translation)	على المدى الطويل يموت	المدخنين لمدة طويلة يتعرضون للوفاة
ميزانية للأمم المتحدة أكثر قوة وكفاءة وفعالية (Economic Translation)	a budget for a stronger United Nations and the efficiency and effectiveness	a budget for stronger and more efficient and effective United Nations
بما يوفر أموال دافعي الضرائب الأمريكيين بالملايين (Economic Translation)	providing money U.S. taxpayers millions	provides millions from U.S. taxpayers
درب الانضباط المالي الحقيقي والإصلاح المستمر (Economic Translation)	path of fiscal discipline and reform the real constant	A path of real fiscal discipline and constant reform
Under Muammar Qaddafi's 42-year rule (Media Translation)	بموجب المادة 42 عاما الزعيم الليبي معمر القذافي	في فترة حكم الزعيم الليبي معمر القذافي الذي حكم البلاد لمدة اثنتان واربعون سنة
longtime leader's capture and killing in October (Economic Translation)	زعيم لفترة طويلة، وقتل في اكتوبر تشرين الاول	القاء القبض على الزعيم ومقتله في تشرين الاول
a new Smoke free campaign (Scientific Translation)	حملة جديدة الخالية من الدخان	حملة جديدة للتخلص من التدخين
research and consulting organization YouGov	البحوث والاستشارات يوجوف المنظمة	منظمة البحوث والاستشارات يوكوف

(Scientific Translation)		
1,000 smoking adults (Scientific Translation)	البالغين التدخين 1000	ألف من البالغين المدخنين

Acronyms

Acronyms, initializes or simply abbreviations may seem insignificant in the field of lexicography, yet they are a problem for translators, and could easily derail a smooth translation or interpretation. (Adetola Bankole, October 2006). The researcher has found that MT faces difficulty in translating acronyms as it is shown in the following table.

Table No. (6) Acronyms

Original Acronyms	MT	Human Translation
NTC (Media Translation)	NTC	المجلس الوطني الانتقالي
NHS (Scientific Translation)	NHS	الخدمات الصحية الوطنية

It appears in the table no. (6) that MT gives acronyms of the same source language i.e., it fails to transfer the meaning of acronyms into target language.

Prepositions

Machine translation faces difficulty in translating prepositions and analyzes the input in a wrong way. It is one of the thorny issues in MT. Many verbs and nouns have specific prepositions that associate with them and are part of the cohesion of the text. For example "وافقت الدول الأعضاء بالأمم المتحدة" is translated into "Member States approved the United Nations" which is wrong from semantic point of view, while it should be "Member States in the United Nations approved". "under king Idris" translated into "تحت الملك إدريس" whereas it should be "تحت سيادة الملك إدريس" "by 70,000" translated into "70000 بواسطة" whereas it should be "يقدر بحوالي 7000".

The table below shows prepositions.

Table No. (7) Prepositions

Original sentence	MT	Human Translation
وافقت الدول الأعضاء بالأمم المتحدة (Economic Translation)	Member States approved the United Nations	Member States in the United Nations approved
under king Idris (Media Translation)	تحت الملك إدريس	تحت سيادة الملك إدريس
by 70,000 (Scientific Translation)	70000 بواسطة	يقدر بحوالي 7000

Structural Problems

Word Order

Sentences in Arabic can be either nominal (Subject Verb Object: SVO) or verbal (VSO). Arabic SVO order corresponds to the English word order and translates much more easily than the VSO order. MT will face problems of word order and syntax if

the languages belong to different families. For example, "بينما طالبت البلدان النامية" is translated into "while demanding that developing countries" while it should be "while the developing countries demanded".

"لم يكن احتلال العراق كارثياً فقط على العراق وأهله" Google translates this sentence into "Was not only disastrous occupation of Iraq on Iraq and its people". As it should be "The occupation of Iraq was not only a catastrophe on Iraq and its people". The table below displays word order.

Table No. (8) Word Order

Original sentence	MT	Human Translation
بينما طالبت "البلدان النامية" (Economic Translation)	while demanding that developing countries	while the developing countries demanded
لم يكن احتلال العراق كارثياً فقط على العراق وأهله (General Translation)	Was not only disastrous occupation of Iraq on Iraq and its people	The occupation of Iraq was not only a catastrophe on Iraq and its people

Subject-Verb Agreement

Agreement between verb and subject is one of the problematic issues in MT. The subject and verb must agree in number: both must be singular, or both must be plural. The following table states that agreement causes challenges.

None of the examples below have been translated accurately with Google as they did not make the subject agree in number with their verbs and ill-ordered translation in all of the examples below. Number agreement between a subject and a verb: A singular noun in the subject position regularly occurs with a singular verb (e.g., "the dog runs"), and a plural subject noun regularly co-occurs with a plural verb (e.g., "the dogs run"). The general rule is "A verb must agree with its subject in person and number. The table below displays Verb-Subject.

Table No. (9) Verb-Subject Agreement

Original sentence	MT	Human Translation
وكانت الولايات المتحدة والدول الأوروبية التي تتعرض للارزمة قد ضغطت من أجل خفض الميزانية (Economic Translation)	The United States and European countries exposed to the crisis has pushed for budget cuts	The United States and European countries, which have been exposing to the crisis, have pushed for budget cuts
وكانت الولايات المتحدة قد اتخذت خطاً متشدداً (Economic Translation)	The United States has taken a hard line	The United States have taken a hard line
long-term smokers die (Scientific Translation)	على المدى الطويل يموت	المدخنين لمدة طويلة يتعرضون للموت
More than half of smokers underestimate (Scientific Translation)	اكثر من نصف المدخنين عن التخمين يقلل	اكثر من نصف المدخنين يستخفون
53% of smokers underestimate (Scientific Translation)	% من المدخنين نقلل 53	53% من المدخنين يستخفون

Long sentences

In view of the differences between languages in their systems and structures, it appears that MT can't convey long Arabic sentences into English correctly and accurately. The table below shows this problem.

Table No. (10) Long sentences

Original sentence	MT	Human Translation
فعودة الجنود من العراق حاملين معهم نفسيات محطمة وسلوكاً عدوانياً وعنفاً تدرّبوا عليه ومارسوه على مدار 9 سنوات، أصبح من أخطر ما يهدّد الولايات المتحدة (General Translation)	return of troops from Iraq carrying their broken psyches and aggressive and violent behavior by trained and practiced for over 9 years, it has become the most serious threats to the United States	The American troops' return from Iraq with broken psyches and aggressive and violent behavior which they were trained and practiced for over years has become the most serious threats to the United States

Passive voice

The passive construction in English presents difficulties for translation into Arabic due to the different structure of two languages mentioned above as it is shown in table 11.

Table No. (11) Passive voice

Original sentence	MT	Human Translation
Libya was occupied for decades by various nations. (Media Translation)	احتلت ليبيا على مدى عقود من قبل الدول المختلفة	كانت ليبيا محتلة على مدى عقود من قبل امم مختلفة
Sharkas had been appointed to the same post by Qaddafi (Media Translation)	كان قد عين في منصب شركس نفس القذافي	قذافي عين شركس في نفس المنصب

Hypotheses:

The First hypothesis says: Machine translation can't produce a text or sentence of the same quality as that of a human being. It really appeared through the analysis of the data that MT can't produce a text or sentence as good as that of human being, for example:

Original sentence	MT	Human Translation
وافقت الدول الأعضاء بالأمم المتحدة، وعددها 193 بلداً. (Economic Translation)	Member States approved the United Nations, and the 193 countries	<i>The</i> members States in the United Nations, which <i>counted</i> 193 countries, approved

The Second hypothesis says: Machine translation can't convey the meaning as clear as human being does. MT sometimes can't transfer the meaning like human being does as the following example proved it.

Original sentence	MT	Human Translation
Sharkas had been appointed to the same post by Qaddafi (Media Translation)	كان قد عين في منصب شركس نفس القذافي	قذافي عين شركس في نفس المنصب

The Third hypothesis says: Machine translation will face difficulties unless it has a good system of semantics, morphology and syntax i.e., linguistics of different languages.

The researcher found during the analysis of the data that MT is problematic unless it has a good system of semantics, morphology and syntax as it is shown in the following examples:

Original sentence	MT	Human Translation
لم يكن احتلال العراق كارثياً فقط على العراق وأهله (General Translation)	Was not only disastrous occupation of Iraq on Iraq and its people	The occupation of Iraq was not only disastrous on Iraq and its people
Only the 1969 date of his coup was <i>marked</i> . (Media Translation)	اتسمت فقط لعام 1969 تاريخ انقلابه	يحييون ذكرى تاريخ انقلاب 1969 فقط

The Fourth hypothesis says: Machine translation will never replace human translators.

Due to the different challenges faced by MT, the researcher sees that MT will never replace human translators.

Original sentence	MT	Human Translation
under king Idris (Media Translation)	تحت الملك إدريس	تحت سيادة الملك إدريس
research and consulting organization YouGov (Scientific Translation)	البحوث والاستشارات يوجوف المنظمة	منظمة البحوث والاستشارات يوكوف

The Fifth hypothesis says: Problems will arise among languages belonging to different families.

MT has difficulty in conveying meaning among languages that belong to different families like English (Indo-European Family) and Arabic (Semitic family), for example:

Original sentence	MT	Human Translation
بينما طالبت "البلدان النامية" (Economic Translation)	while demanding that developing countries	while the developing countries demanded
وكانت الولايات المتحدة قد اتخذت خطاً متشدداً (Economic Translation)	The United States has taken a hard line	The United States have taken a hard line

The Sixth hypothesis says: Machine translation will face challenges in translating multi word forms as in the examples below:

Original Collocation	MT	Human translation
<i>long-term smokers</i> (Scientific Translation)	طويلة الأجل المدخنين	المدخنين لامد طويل

درب الانضباط المالي الحقيقي والإصلاح المستمر (Economic Translation)	path of fiscal discipline and reform the real constant	A path of real fiscal discipline and constant reform
--	--	--

The Seventh hypothesis says: To a large extend machine translations can't convey acronyms meanings from SL to TL as shown below.

Original Acronyms	MT	Human Translation
<i>NTC</i> (Media Translation)	NTC	المجلس الوطني الانتقالي
<i>NHS</i> (Scientific Translation)	NHS	الخدمات الصحية الوطنية

5. Conclusion

Currently machine translation (Google Translation) is probably the most important field of computational linguistics. Machine translation has begun to become an indispensable tool for many users, and its significance is dramatically increasing day by day. MT has been given more interest recently due to different reasons such as: information revolution, advance in telecommunication and other fields of everyday life, huge amount of data that need to be translated in time that it is impossible to get human translator or the shortage in the number of qualified translators.

The obstacles to translating by means of the computer are primarily linguistic. To overcome them it is necessary to resolve the ambiguities that pervade a natural language when words and sentences are viewed in isolation. Texts are normally not ambiguous to their readers; the challenge for researchers is to develop computer programs with something like the ability of the ordinary human reader to extract the intended meaning from the string of words that forms a text. To put the matter another way, the computer must be taught to understand the text- a problem in artificial intelligence. (Lehrberger & Bourbeau, 1984).

As laid out at the beginning of this paper, the objective of this research is to investigate problems and challenges faced by machine translation (Google), mainly lexical and structural ones from and into English and to help scientists to develop MT systems. Based on the analysis of the obtained data in this study, the researcher has come up with the following conclusions:

1. Although MT has been progressed and given attention recently by scientists and programmers, there is still a long way to go forward as it is still defective in comparison with the human translation and there is a lot that we can do to improve the quality of MT output and increase its usefulness
2. MT (Google) sometimes translates texts that can't be understandable unless one has previous knowledge on that text and it has not been able to deliver fully automated high-quality translations.
3. MT is not accurate.
4. It has been proved that MT has lexical and structural problems related to: deletion, Non-vocalization, homographs, collocations, acronyms, wrong word order, subject-verb agreement, passive voice ...etc.
5. Some words either remained untranslated, or deleted in MT (Google System).

References

Alzebaree, Y., Yavuz, M. A. (2017). Realization of the Speech Acts of Request and Apology by Middle Eastern EFL Learners. Eurasia Journal of Mathematics, Science and Technology Education, 13(11), 7313-7327. <https://doi.org/10.12973/ejmste/79603>.

- Achiba, M. (2003). Learning to request in a second language. Clevedon, UK: Multilingual Matters.
- Aronin, L., & Singleton, D. (2012). Multilingualism. Amsterdam: John Benjamins. Austin, J. L (1965). How to do things with words. New York: Oxford
- Achiba, M. (2003). Learning to request in a second language. Clevedon, UK: Multilingual Matters.
- Aronin, L., & Singleton, D. (2012). Multilingualism. Amsterdam: John Benjamins. Austin, J. L (1965). How to do things with words. New York: Oxford
- Abdel, A. (2002). Implication of the agreement features in machine translation. (Master's thesis) Retrieved on 08 October 2011 from: www.attiaspace.com/Publications%5CDissertation.pdf.
- Anastasiou. (2010). Idiom Treatment Experiments in Machine Translation. Cambridge Scholars Publishing. Retrieved on 01 November 2011 from www.c-s-p.org/Flyers/978-1-4438-2515-3-sample.pdf.
- Arnold, Balkan, Humphreys, Meijer & Sadler, (2001, March). Machine Translation: An Introduction Guide. Retrieved on 06 October 2011 from: www.mt-archive.info/Arnold-1994-0.pdf.
- Bankole, October 2006. Dealing with Abbreviations in Translation. From Translation Journal, Volume 10, No. 4. Retrieved on 05 November 2011 <http://www.bokorlang.com/journal/38acronyms.htm>.
- Bessou & Touahria, 2011. Morphological Analysis and Generation for Machine Translation from and to Arabic. International Journal of Computer Applications (0975-8887) Volume 18- No.2. Retrieved on 25 October 2011 from: www.ijcaonline.org/volume18/.../pxc3872896_pxc3872896.
- Boussalia, (2010). Students' Difficulties in English- Arabic Translation of Collocations.
- Corr, (2005). Identification and Avoidance of Critical Ambiguity. Retrieved on 20 October 2011 from www.scss.tcd.ie/undergraduate/bacsl/bacsl_web/corrms0405.pdf.
- Cuyckens & Zawada, (2001). Polysemy in Cognitive Linguistics. Retrieved on 29 October 2011 from <http://books.google.com/books?id=Bt2DA6qj1xIC&printsec=frontcover&dq=Polysemy+in+Cognitive+Linguistics>.
- Daxbock, A. (2010). Machine Translation: History - Approaches – Evaluation.
- Guerra, (2000). Machine Translation Capabilities and Limitations. Lengua Inglesa. Universitat De Vslencia. Retrieved 10 October 2011 from <http://books.google.com/books?id=7TE3avRZiSoC&pg=PT1&lpg=PT1&dq=Machine+Translation+Capabilities+and+Limitations.+Lengua+Inglesa>.
- Hamdalla, (1998, Sep.) Problems and Approaches to Translation with Special Reference to Arabic. J. King Saud Uni., Vol.10, Lang. & Transl., p.24).
- Hatim and Munday, (2004). Translation: An advanced resource book. Retrieved 01 September 2011 from http://books.google.com/books?id=3gZc6rCduLYC&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false.
- Hoppe, (June 2005). Machine Translation Why Is It So Difficult. Mid-Sweden University. Retrieved 29 October 2011 from http://www.google.com/#hl=en&cp=65&gs_id=3&xhr=t&q=Machine+Translation+Why+Is+It+So+Difficult.+Mid-Sweden+University&pf=p&scli.
- Hui, A (1998). The Role of Morphology in Machine Translation. Canada, University of Waterloo. Retrieved on 10 October 2011 from: www.cs.uwaterloo.ca/research/tr/1998/22/CS-98-22.ps.Z.
- Hutchins, A. (1995). Machine Translation: A Brief History. Retrieved on 05 October 2011 from: www.infiamerica.org/documentos_pdf/bar05.pdf. Oxford: Pergamon Press, 1995. Pages 431-445.
- Izwaini (n.d., p. 119-135). Problems of Arabic Machine Translation: evaluation of three systems. Retrieved on 03 October 2011 from: www.mt-archive.info/BCS-2006-Izwaini.pdf.
- Izwaini (n.d., p. 120). Problems of Arabic Machine Translation: evaluation of three systems. Retrieved on 01 November 2011 from: www.mt-archive.info/BCS-2006-Izwaini.pdf.

- Karamat, 2006. Verb transfer for English to Urdu machine translation. Master of Science (Computer Science). National University of Computer & Emerging Sciences. Retrieved 10 November 2011 from http://www.google.com/#sclient=psy-ab&hl=en&source=hp&q=Verb+transfer+for+English+to+Urdu+machine+translation&pbx=1&oq=Verb+transfer+for+English+to+Urdu+machine+translation&aq=f&aqi=&aql=&gs_sm=s&gs_upl=399923139992313140164811110101010111791117917-11110&bav=on.2,or.r_gc.r_pw..cf.osb&fp=1e0e72d07edee1f3&biw=916&bih=619
- Kumar. (1994). Machine Translation. Desidoc, Metcalfe House, Delhi- 110 054. Retrieved 20 October 2011 from publications.drdo.gov.in/gsd/collect/dbit/index/.../dbit1406003.pdf.
- Lehrberger&Bourbeau, (1984). Machine translation: linguistic characteristics of MT systems and general methodology of evaluation. Philadelphia: John Benjamins Publishing Co.. Retrieved 01 September 2011 from http://books.google.com/books?id=YUNLlurHNAEC&printsec=frontcover&source=gbp_ge_summary_r&cad=0#v=onepage&q&f=false.
- McCarthy, (2002). An Introduction to English Morphology. Edinburgh University Press Ltd.
- Mohamed, A. (2000). Machine translation of noun phrases: from English to Arabic. (Master's thesis) retrieved on 28 October 2011 from: www.arc.sci.eg/narims_upload/claesfiles/3805.pdf.
- Murphy, (2010). Lexical Meaning. Retrieved on 27 October 2011 from <http://books.google.com/books?id=o3g5dMofckgC&printsec=frontcover&dq=Lexical+Meaning>.
- Schulze, (2001). The identification of word classes in connection with the differentiation between homonymy and polysemy. Retrieved on 30 October 2011 from http://books.google.com.pe/books?id=wSCjd0xX7rgC&printsec=frontcover&source=gbp_ge_summary_r&cad=0#v=onepage&q&f=false.
- Şenkal, 2003. An Approach for Machine Translation between Turkish and Spanish. Master of Science in Computer Engineering. Boğaziçi University. Retrieved 01 November 2011 from www.cmpe.boun.edu.tr/graduate/allthesis/m_Senkal_Metin.pdf.
- Strazny, (2005). Encyclopedia of Linguistics. An Imprint of the Taylor & Francis Group, New York. Oxon.
- Trask, (2007). Language and linguistics. Taylor& Francis Group.