MACHINE TRANSLATION IN SAUDI ARABIA

by

Faten AlMutawa

A Thesis Presented to the Faculty of the
American University of Sharjah
College of Arts and Sciences
in Partial Fulfillment
of the Requirements
for the Degree of

Master of Arts in
Translation and Interpreting (English/Arabic/English) (MATI)

Sharjah, United Arab Emirates

(December 2012)

Approval Signatures

We, the undersigned, approve the Master's Thesis of [*candidate's full name].

Thesis Title:

Signature **Date of Signature**
(dd/mm/yyyy)

_____ _____
[Name]
[Academic title]
Thesis Advisor

_____ _____
[Name]
[Academic title]
Thesis Committee Member

_____ _____
[Name]
[Academic title]
Thesis Committee Member

_____ _____
[Name]
[Program/Department, Coordinator/Director/Head]

_____ _____
[Name]
[CAS/CEN/CAAD Graduate Programs Director]

_____ _____
[Name]
Dean of College/School

_____ _____
Dr. Khaled Assaleh
Director of Graduate Studies

Acknowledgments

This thesis would never have been finished without the guidance of my supervisor, Dr. Sattar Izwaini, and the help and support of my family. My sincerest gratitude to my supervisor for the support, patience, and knowledge he showed throughout the stages of research and writing of this thesis. I am deeply grateful to my parents and my brothers for their help and support.

I would like to extend my thanks and appreciation to my other professors, Dr. Basil Hatim, Dr. Said Faiq, and Dr. Ahmed Ali, who provided their knowledge and experience during my graduate studies in the MATI program.

I would also like to thank Shura Council staff, especially Mr. Salih Saber, for their cooperation and providing me with helpful information.

Grateful acknowledgment is extended to ATA Software Technology Limited; John Benjaimns Publishing Company; Palgrave Macmillan; Routledge, Taylor and Francis Group; Science Application International Corporation, and Springer Science and Business Media B.V. for granting their kind permission for the use of screenshots and figures.

Abstract

Recently, there has been an increasing interest in machine translation technology in the Arab world and the world at large, especially with the growing demand for translation. The aim of this thesis is to investigate how widely Arabic machine translation systems are used and researched in the Arab world and what can be done to achieve real progress in this field. A survey was carried out in Saudi Arabia to collect the necessary data. Two questionnaires were used to assess how many Arabic machine translation systems are used by Saudi organizations, how much interest exists in machine translation in Saudi universities and research centers, and how much research is being conducted on issues of machine translation. The findings of this study indicate that most Saudi organizations and translation agencies do not trust or are not interested in machine translation. Only a few universities have conducted research in this field. It is recommended that more attention be paid and more research be conducted to get the most use out of this technology and that more efficient Arabic machine translation systems be developed.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Technology and machines are ubiquitous in daily life as they are used in all fields, even in tasks usually carried out by humans such as translation. With the increasing need for translation, improving machine translation (MT) technology has become a necessity. Research on MT has been conducted since the 1950s. Moreover, MT has proven to be a big aid to professionals. Information technology applications, including translation software, is classified as one of the translation aids belonging to the domain of applied translation studies research (Holmes, 2000, p. 173; see also chapter 2 of this paper). Research has led to many MT systems that have been developed and used in many countries. Some of these systems include the Arabic language and are used in both Western and Arab countries.

The aim of this study is to provide a scholarly account of MT in Saudi Arabia and to discover to what extent MT has been established in that country both as an area of research and as a professional practice. A further goal is to explore what measures might be taken to achieve substantial progress in this field, based on such questions as do Saudi organizations benefit from this technology, and do they use the available Arabic MT systems? Are Saudi universities and their faculties interested in this field? Are linguists, translators, computer engineers, and software programmers interested in MT? And if there is any research in this field, is it adequately funded by the Saudi government or universities?

To discover whether Saudi governmental and nongovernmental organizations are interested in MT or use it, a questionnaire was designed and submitted to them for their consideration (as described in chapter 4). As translation agencies play an important role in the translation industry, data were collected from a number of agencies to ascertain their views regarding this technology and also whether their business derives any benefit from it. Another questionnaire was designed for Saudi universities that offer degrees in translation to ascertain to what extent they are interested in teaching MT and conducting research on MT. Data were also collected from Saudi universities that offer degrees in computer science and information technology to find out whether research is being

conducted on MT and related fields such as natural language processing (NLP) and Arabic computing (i.e., Arabization of computers).

This thesis is divided into five chapters: Chapter 1 is an introduction to the study. Chapter 2 provides a review of translation studies and identifies the place of the present thesis within the body of research of this discipline. It also discusses MT and related areas such as Human Aided Machine Translation (HAMT) and Machine Aided Human Translation (MAHT). It introduces the two types of MT: bilingual and multilingual. Chapter 2 also introduces basic approaches to MT and discusses the different uses of MT such as dissemination, assimilation, interchange, and data access. It discusses which text types can be translated by MT, as well as the benefits and drawbacks of MT, its limitations, and the quality of its output.

Chapter 3 discusses the importance of Arabic MT. It explains special features of the Arabic language, which can cause problems and become challenging for MT. It introduces the companies doing research on and producing Arabic MT systems and describes Arabic MT and online translation systems currently available.

Chapter 4 discusses the survey carried out on Saudi organizations, translation agencies, and universities. Then the findings of the survey are analyzed and commented on.

Chapter 5 concludes the thesis by summarizing its findings and providing recommendations.

Chapter 2: Machine Translation

This chapter sets the scene with a definition of MT and discusses the importance of this technology. The different approaches to MT will be introduced. Its uses, benefits, and drawbacks will also be explored.

Translation Studies and Holmes's Map

Translation studies refers to the study of translation as an academic subject. Holmes (2000) describes this discipline as being concerned with "the complex of problems clustered around the phenomenon of translating and translations" (p. 173). He proposes an overall framework, describing what translation studies cover, and divides it into pure and applied areas (see Figure 1):

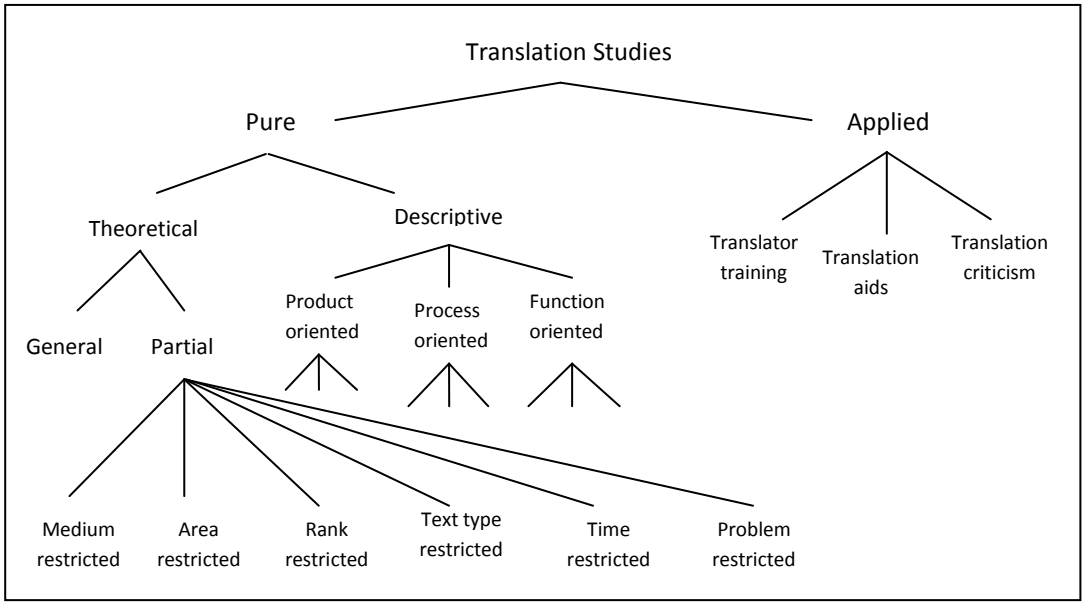

Figure 1. Holmes's basic map of translation studies. From *Descriptive Translation Studies and Beyond* (p. 10), by G. Toury, 1995. Copyright 1995 by John Benjamins Publishing. Reprinted with permission.

According to Munday (2001, pp. 10-12), pure areas focus on research in the description of the phenomenon of translation (descriptive translation theory) and the

establishment of general principles to explain and predict such a phenomenon (translation theory).

- The descriptive branch has three possible foci: product, function, and process.
- The theoretical branch is divided into general and partial theories. General theoretical studies are concerned with every type of translation, whereas partial theoretical studies are restricted according to certain parameters: medium, area, rank, text type, time, and problem.

Munday (2001, p. 12) explains that applied areas (see Figure 2) focus on research in:

- Training of translators, which includes teaching methods, testing techniques, and information design.
- Translation aids, which includes dictionaries, grammars, and information technology (IT) applications such as translation software, online databases, and use of the Internet.
- Translation criticism, which includes marking of students' translations and reviewing published translations.

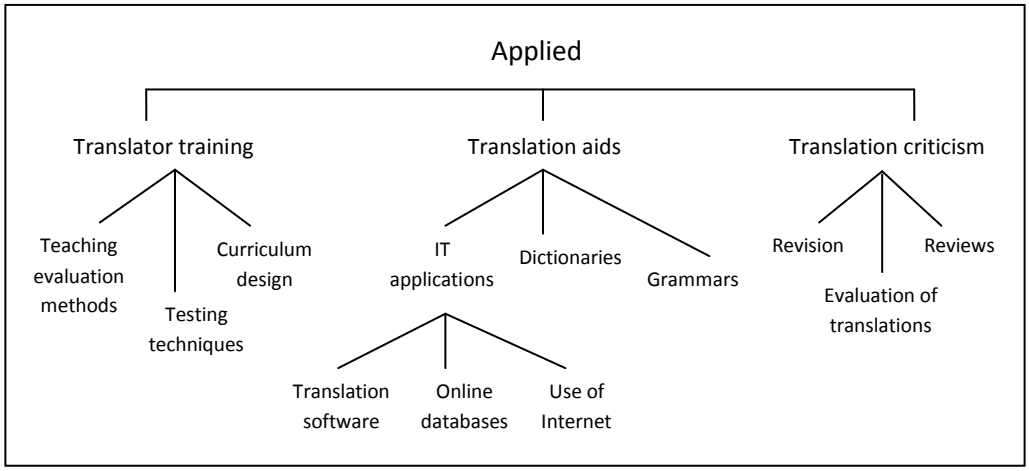

Figure 2. The applied branch of translation studies in Holmes's map. From *Introducing Translation Studies: Theories and Applications* (p. 13), by J. Munday, 2001. Copyright 2001 by Routledge. Reprinted with permission.

Figure 2 shows that MT as an IT application falls under translation aids within the applied areas of translation studies.

What is Machine Translation (MT)?

The European Association for Machine Translation (EMAT, n.d.) defines MT on its Website as "the application of computers to the task of translating texts from one natural language to another." According to Quah (2006), MT systems "are purely automatic with no human intervention during the actual translation process" (p. 173).

Although the term *MT* refers only to automated translation without human intervention, there are other terms such as *MAHT* and *HAMT* that are worth highlighting. MAHT, which is also referred to as Computer Aided Translation (CAT), is defined by Lehrberger and Bourbeau (1988) as "basically human translation with only limited assistance from the machine" (p. 6). HAMT, on the other hand, is a process where "the system itself takes the main responsibility for translation, with human assistance to help in the process when needed" (Hutchins & Somers, 1992, p. 150).

Extensive research has been conducted on MT since the 1950s in the USA, Europe, and Japan. Hutchins (2002, pp. 159-160) explains that, at the beginning, the assumed goal of MT was to translate all kinds of documents with high quality. However, as it turned out, that goal was impossible to achieve; therefore, it was amended such that MT output was to be revised if the translation was to be published. At the same time, it was found that a rough MT output could be very useful for assimilation and information gathering; however, this use was largely ignored.

For many years, HAMT was used by many large organizations such as the European Union with large volumes of technical and administrative documents to be translated into different languages. Since the revision of MT output was expensive, controlled language was also used to reduce the cost of the revision. A controlled language aims to reduce lexical ambiguity and simplify complex sentence structures to enhance the comprehensibility of the original texts by applying certain rules such as using only approved terminology and sense; avoiding ambiguous words, pronouns, and phrasal verbs; and using articles and short sentences (Hutchins, 2005). At the same time,

human translators have been assisted by CAT (or HAMT) tools such as dictionaries, glossaries, terminology management, and later on translation memories (Hutchins, 2005).

Hutchins (2002, pp. 159-160) also argues that MT output does not always have to be of publishable quality. Speed and accessibility may, sometimes, be more important. From the beginning, it was found that a rough MT output could be used for translating technical reports, administrative memoranda, personal correspondences, or any other documents that were to be read by just one or two people. Moreover, with the increasing use of personal computers and the Internet since the 1990s, the use of rough MT output has significantly expanded.

According to Hutchins (2003, p. 181), much progress has been achieved with MT since the 1980s. Obviously, MT systems are much faster, text input is much easier to deal with, and the output is much easier to read than it had been in the past, but more important is that, in the words of Hutchins, "more and more large organizations can and do save translation costs by using MT systems as aids in the production of draft translations—which was almost inconceivable twenty years ago" (p. 181). Hutchins and Somers (1992) reflect,

> There may still be many misconceptions about what has been achieved and what may be possible in the future, but the healthy state of MT is reflected in the multiplicity of system types and of research designs which are now being explored, many undreamt of when MT was first proposed in the 1940s. (p. 9)

Types of MT

According to Hutchins and Somers (1992, pp. 69-70), two types of MT are extant: bilingual and multilingual. They can be categorized either according to the translation direction, that is, unidirectional systems (only in one direction); bidirectional systems (in both directions); or according to the number of languages involved, that is, bilingual systems, which translate between a single pair of languages, or multilingual systems, which translate between more than two languages.

Multilingual systems, according to Hutchins and Somers (1992, pp. 69-70), can be organized as follows:

- Complex systems that translate between a large number of languages in every combination such as the European Commission's Eurotra project, which translates between all nine languages of the European communities in all directions.
- More modest systems that translate from English into three other languages in one direction.
- Intermediate systems that translate between five languages (e.g., Japanese and four European languages), but not in all combinations of pairs and directions; thus, it may translate from Japanese into the four European languages or from these four languages into Japanese, but not between the four European languages.

Basic Approaches to MT

Three generations of MT systems are divided, as explained by Quah (2006, p. 68), into three kinds of systems: direct, rule-based, and corpus-based (see Figure 3). These generations will be discussed further in subsequent sections.

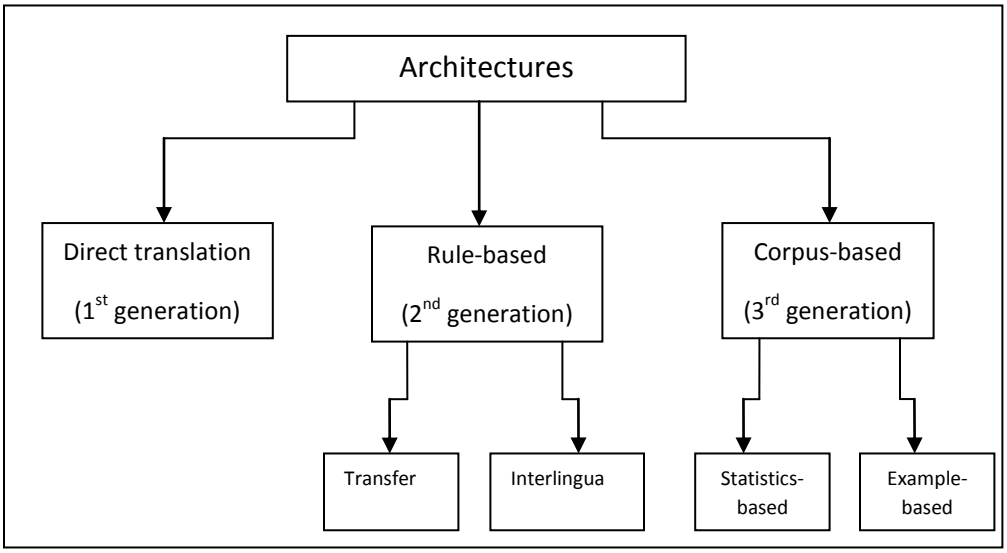

Figure 3. MT architectures. From *Translation and Technology* (p. 68), by C. Quah, 2006. Copyright 2006 by Palgrave Macmillan. Reprinted with permission.

Direct approach (first generation). The basic component of this approach is the dictionary, which provides all lexical and syntactic information and the translation in the

target language (TL) for each entry in the source language (SL). When translating, the system starts with replacing the words in the SL with their corresponding words in the TL. Then, the words are rearranged to produce a text in the TL. This approach can be used to translate between two closely related languages in one direction (Vauquois, 2003, pp. 333-334).

One example of this approach is Systran (as cited on Global Security Website, n.d.), which is one of the oldest MT systems and is still used today. According to Global Security (n.d.), Systran's first version was published in 1969. Over the years, the system has been developed, but still its translation is mainly based on very large bilingual dictionaries (as described in chapter 3). This approach does not use parsing principles but depends instead on well-developed dictionaries, morphological analysis, and text processing software (Global Security, n.d.).

Rule-Based MT systems (second generation). This approach works on a sentence-by-sentence basis. It starts with analyzing the words individually, identifying the part of speech of each word and its possible meanings, using a dictionary (Somers, 2011, p. 433). If the words cannot be found in the dictionary because of a language's rich systems of inflection and derivation, morphology rules can be used to analyze the sentences and suggest possible translations.

According to Carl and Way (2003), "Rule-Based Machine Translation (RBMT) is characterized by linguistic rules used in translation" (p. xviii). RBMT systems consist of a series of processes that analyze input texts morphologically, syntactically, and semantically and follow this with a process of generating text as a result of a series of structural conversions based on an internal structure. These processes are controlled by a dictionary and a grammar, developed by linguists (Carl & Way, 2003, p. xviii).

Transfer approach. This approach is an RBMT approach and consists of three stages. The first stage is the analysis stage in which the SL text is converted into an abstract SL representation. The second stage is a transfer, which takes place at both lexical and structural levels, generating corresponding structures in the TL. The third

stage is the generation of the translation. Three dictionaries are needed for this approach: an SL dictionary, a bilingual dictionary, and a TL dictionary (Tucker, 1987, p. 23).

According to Kit, Pan, and Webster (2002), "the transfer approach incorporates language analysis and representation at various linguistic levels, but cannot find adequate knowledge to resolve ambiguities involved in the language analysis, transfer, and generation" (p. 57).

Interlingua approach. Using this approach, translation is done via an intermediary (semantic) representation of the source text, which is called Interlingua (IL). IL is a language-independent representation of the SL text from which translations can be generated to different TLs. This approach assumes that it can convert source texts into representations common to more than one language. Then, from such IL representations, texts are generated in other languages. So, translation is done in two stages: the first stage proceeds from the SL to the IL; the second stage proceeds from the IL to the TL (Global Security, n.d.). When adding a new language to the system, only an analysis grammar and a generation grammar module need to be added (Hutchins & Somers, 1992, p. 74).

Corpus-Based MT systems (third generation). Corpus-Based MT (CBMT) systems use a bilingual parallel corpus (a database of translated sentences) to derive the required knowledge for new translations. It is an alternative approach that has been proposed to overcome the knowledge acquisition problem of Rule-Based MT. Within the Corpus-Based MT approach, two main directions can be distinguished: Example-Based MT and Statistics-Based MT (Carl & Way, 2003, p. xviii).

Example-Based approach. Arnold, Balkan, Meijer, Humphreys, and Sadler (1994) explain this approach as aiming to "collect a bilingual corpus of translation pairs which use a match algorithm to find the closest example to the sentences of the source text. This process produces a translation template which can be then completed by word-for-word translation" (p. 188).

The ideal translation unit in an Example-Based MT system is the sentence. When a translation of an identical sentence is not available in the bilingual corpus, the system uses similarity metrics to find the best matching translation examples. It replaces,

substitutes, modifies, or adapts suitable subsequences to generate the translation (Carl & Way, 2003, p. xix).

Statistics-Based approach. This approach is based on the idea that a computer program can learn how to translate by analyzing huge amounts of data from previous translations and then assessing statistical probabilities to decide how to translate a new input (Somers, 2011, p. 434). Carl and Way (2003) explain that a Statistics-Based MT system learns a translation model from a bilingual parallel corpus and a language model from a monolingual corpus. So, when translating, it searches for the best translation by maximizing the probability according to the two models (p. xix). This approach depends on numerical knowledge, which is extracted from a corpus, and, while it reduces the need for human knowledge, search costs are high (Chen & Chen, 1996, p. 160).

Other Approaches

Knowledge-Based approach. This approach displays extensive semantic and pragmatic knowledge of a certain domain. Nirenburg (1989, p. 344) argues that even this approach is useful in very small domains, but with paying more attention to the acquisition of large knowledge bases and with the advent of new tools (including representation languages, human-computer interfaces, database management systems, etc.), its practicality can be increased. Knowledge-Based MT systems are based on taxonomies of knowledge and IL representations, but it is hard to build a knowledge hierarchy and to represent knowledge (Chen & Chen, 1996, p. 160).

Hybrid approach. A hybrid approach, as the term implies, combines two different MT systems. In this way, the same data structures are shared by the different components of the two systems. Yet, some of these components may modify or adjust certain processing resources of another component to improve translation quality. For example, when combining a statistics-based system and a rule-based system, the statistical data are added to the lexical resources of the rule-based system, thus judging different translation candidates for a given textual or thematic context and, therefore, producing a better translation (Schäler, Way, & Carl, 2003, p. 102).

Interactive approach. With an interactive MT system, a translator has control over the translation process and the output. This approach is different from that of HAMT, in which the machine pauses and asks the user to solve the lexical or syntactic ambiguity. However, they are similar in dealing with the input (Quah, 2006, p. 85). The drawback with this approach is that there are sometimes many interactions for a single sentence, or the same interactions with the same lexical or syntactic problems may be repeated for different sentences (Hutchins & Somers, 1992, pp. 153-154).

Uses of MT

Translation is mainly used for one of the four following functions and so is MT (Hutchins, 2005, p. 7):

1. Dissemination: Texts to be translated for dissemination need to be of high quality. So, if MT is used to translate such texts, then human assistance is necessary, whether for pre-editing the input, post-editing the output, using a controlled language, or restricting the system to a specific subject domain.

2. Assimilation: When the texts to be translated will be used for monitoring or filtering information, or if the recipients only need to get a general idea about what the text conveys, then the MT will do, and there is no need for a good quality translation.

3. Interchange: When the translation is needed for communication between individuals who speak different languages, by correspondence, e-mail, or telephone, then again any translation will do as long as the communicators understand the message they receive and it conveys their intentions.

4. Database access: Nowadays, many people use translations, even rough ones, for searching the Internet and Websites and for accessing databases to get information in foreign languages.

MT has also been used in large companies and as software on personal computers. However, MT systems have been developed to be used in hand-held devices (pocket translators used by travelers, mobiles) as well, and on the Internet as AltaVista's Babelfish service, Gist-in-Time, ProMT, PARS, and many others, and for localization in e-mails, chat rooms, and social networking sites (Hutchins, 2011, pp. 441-446).

According to Hutchins (2005, p. 16), MT is used for the following purposes:

- Document drafting (in poorly known languages).
- By tourists and shoppers: so far only dictionaries of words and phrases (using hand-held devices).
- Scanner-translator (scan/OCR [optical character recognition] / MT/ print) - desktop, portable, or online.
- Interchange with deaf and hearing impaired: translation into sign languages [mainly research so far].
- Information retrieval (IR): translation of search terms [very active field].
- Information filtering (intelligence):
  - for human analysis of foreign language texts.
  - document detection (texts of interest); triage (ranking in order of interest).
- Information extraction: retrieving specific items of information (e.g., news analysts).
- Summarization: producing summaries of foreign language texts.
- Multilingual generation from (structured) databases.
- Television subtitling.

Benefits of MT

MT has many benefits such as the following (Dilmanc, n.d.; The Language Translation, n.d.):

1- Productivity: It improves the productivity of human translators who are willing to perform pre- or post-edit translations or both.

2- Speed: It is much faster than a human translator.

3- Low cost: It is a one-time cost, namely, the cost of the tool and its installation.

4- Confidentiality: People can use it to translate private e-mails or financial documents.

5- Consistency: It keeps translation consistent. So, there is no need to go back to previous pages to see how a certain word was translated.

6- Universality: Human translators usually specialize in certain domains, whereas MT can translate texts in different domains. One only needs to switch on a corresponding setting.

7- Availability: MT is available around the clock.

Drawbacks and Limitations of MT

1. Lack of superior exactness: If accurate translation of official documents, agreements, and so on is needed, then MT cannot be used. However, if it is used, then the output should be pre- or post-edited or both (Dilmanc , n.d.).

2. Inferior translation quality of texts with ambiguous words and sentences: The SL text to be translated by an MT system should be clear and straightforward. If it includes ambiguous words or complex syntax, then a poor translation ought to be expected (Dilmanc, n.d.).

3. MT is based on formal and systematic rules; so, sometimes it will not be able to resolve ambiguity by concentrating on the context and using experience or mental outlook, as human translators are able to do (Dilmanc, n.d.).

4. MT cannot correctly translate expressions (multiword terms) that convey ideas, idiomatic phrases, ambiguous words, complex structures, proverbs, opaque sentences, ellipses, colloquial phrases, and culture-specific aspects (Hutchins, 2011, pp. 445-446).

5. MT cannot retain aspects of source culture or adapt to the target culture. It cannot maintain the same register as the source document, or coin translations for new technical terms (Melby, 1987, p. 145).

6. Machines break down, and replacement can be costly.

Text Types and MT

Almberg (2002) argues that, “indeed, we are all aware today that MT has come to stay—if not for poetry, then at least for domain-specific and repetitive or standardized materials” (p. 187). The following text types, according to Melby (1987, pp. 145-146) and The Language Translation Website (n.d.), can be translated by MT:

- Technical texts: These deal with concrete subjects, follow a regular procedure, have short sentences, and are usually informative texts for which it is not required to understand the content to translate the text.
- Weather forecasts: These have been successfully translated by MT.
- Web pages: MT works well in dealing with the Internet.
- Documents for which the readers need rough translations or which are used in multinational companies.
- Texts in which there are many repetitions.
- Texts of which quick translations are needed or where it is important to maintain consistency.

The following text types, according to Melby (1987, pp. 145-146), are difficult to be handled by MT:

- Literary texts: They usually are difficult to be translated, even for a human translator because creativity and style are very important in this field.
- Legal texts: Their syntax is usually complex, equivalents should be assigned to larger units than just words or phrases, and differences in legal systems should be taken into consideration; therefore, their translations should be accurate.
- Advertising material: Requires creative and catchy translations, which MT cannot produce.
- Texts that deal with cultural aspects or use idiomatic language.

Schäler, Way, and Carl (2003, pp. 104-105) assign three levels to translation quality: high, medium, and low, corresponding to three different types of texts (see Figure 4). At the top, a human translator is chosen to translate texts of a creative nature such as advertisements and literature. At the middle, a human translator and the machine both need to be used to translate a large amount of subject-specific texts such as laboratory reports and manuals because accuracy and consistency are important for such texts. At the bottom, a machine can be used to produce rough translations such as online MT systems of Web pages, services, and general information (Quah, 2006, pp. 154-155).

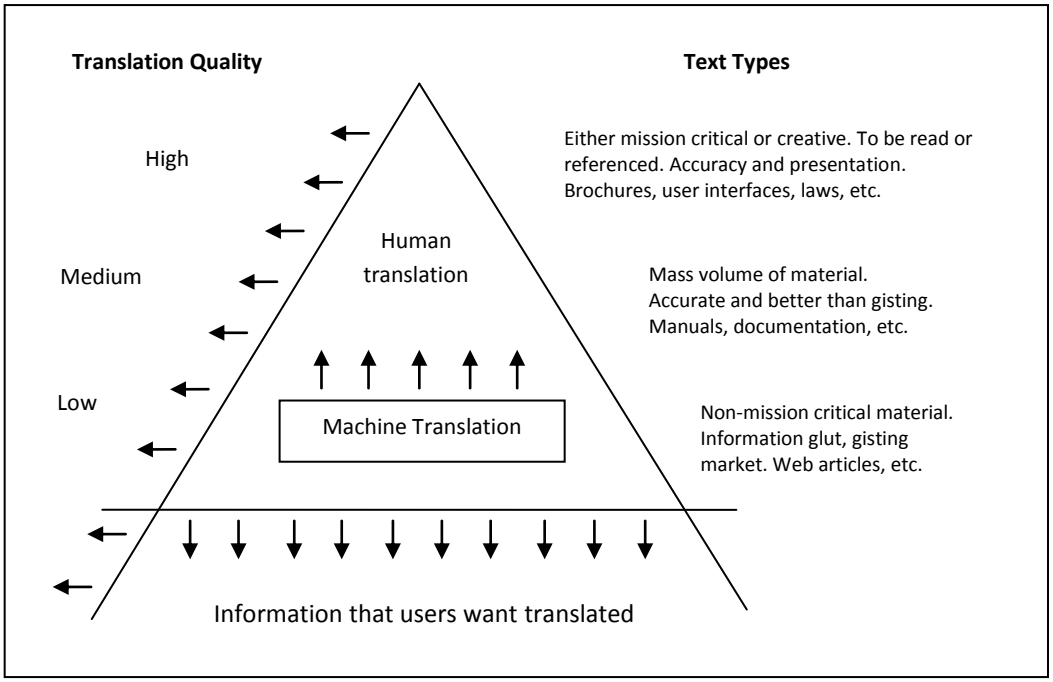

Figure 4. : Schäler, Way, and Carl's model for future use of translation technology. From Example-Based Machine Translation in a Controlled Environment (p. 104). In M. Carl and A. Way (Eds.), *Recent Advances in Example-Based Machine Translation*, 2003. Copyright 2003 by Kluwer Academic Publishers. Reprinted with permission.

The pyramid in this model is expanding horizontally and vertically, driven by two factors: the growing demand for translated material (horizontal expansion) and the increasing availability and accessibility of information in different languages to users of the Web (vertical expansion). At the same time, research and development in translation technology will allow MT to push its way up the pyramid and produce even higher quality translations (Schäler et al., 2003, p. 105).

Improving the Quality of MT Output

MT output can be improved by the following processes: pre-editing the input or post-editing the output or both.

Pre-Editing. Pre-editing the input means avoiding idiomatic expressions, avoiding relative pronouns, breaking up long sentences, using the typical word order of subject—verb—object, using active instead of passive voice, and not splitting phrasal verbs (e.g., look up). Pre-editing also involves using consistent and unified terminology and identifying text that is to remain untranslated such as proper names or addresses (Raído & Austermühl, 2003, p. 247).

Post-editing. Post-editing the output means revising and correcting the output, that is, the raw translation generated by the MT system. Whether post-editing is conducted, and to what extent, depends on the quality of the translation required by the user (Raído & Austermühl, 2003, p. 247).

It can be either rapid post-editing or full post-editing. Rapid post-editing is used for translations intended only for information purposes such as working papers for internal meetings, minutes of meetings, technical reports, or annexes. It is done in order to remove significant errors and thus stylistic issues are not considered. It is also called minimal post-editing because its aim is to provide the minimum amount of necessary corrective work in order for the text to be comprehensible. On the other hand, full post-editing is used for translations intended to be published. It is also called maximal post-editing because its aim is to provide high quality translations. Although using MT can save time, applying full post-editing can be even more time-consuming than translating a text from scratch (Allen, 2003, pp. 302-306).

According to Wilss (as cited by Raído & Austermühl, 2003), "the tasks of pre-editing and post-editing should be put in the hands of professional translators who are aware of the limitations of MT, since without such professional intervention the process is very time-consuming and inefficient" (p. 247).

Summary

In this chapter, MT, its types, approaches, uses, benefits, and drawbacks were explained. In chapter 3, the importance of Arabic MT will be discussed. The different MT systems available for the Arabic language will be introduced.

Chapter 3: Arabic Machine Translation

In this chapter, the importance of Arabic MT for the translation business and for information dissemination is discussed. The challenges for Arabic MT will also be touched upon. Lastly, the different available Arabic MT systems will be outlined.

The Importance of Arabic MT

The revolution of information technology and the dramatic developments that have been initiated by it have increased the need to break through language barriers and to communicate with the world. Therefore, the demand for translation has increased sharply, and due to the large amount of material needing to be translated in every field, MT systems are now used widely in the West (Zantout & Guessoum, 2000; Zughoul & Abu-Ashaar, 2005).

As the Arab world looks forward to improving its quality of life, organizations and individuals need to gain access to new information, discoveries, and technologies in the world and to benefit from others' experiences. This means that more translations will need to be done, and since translators will not be able to cope with the volume of material, there is a place for the use of MT, which can save time and energy (as detailed in chapter 2), at least, when there is merely a need for the gist of a text rather than for a full and accurate translation or when translating Websites and online information. MT can even be used to produce rough translations, with translators post-editing the output, especially when only a rapid post-editing is required. In this way, human translators can save time that they would otherwise spend translating simple or repetitive texts. Zantout and Guessoum (2000) state that "it is obvious, that MT will boost the technology transfer efforts to make more information about new technologies available to Arabs in their native language" (p. 123).

According to Farghaly and Shaalan (2009), Zughoul and Abu-Ashaar (2005), and Zantout and Guessoum (2000), using MT in the Arab world can be helpful in many ways; however, the following benefits currently seem to exist more in theory than in practice:

- To cope with the increasing demand from multinational companies and governments for translation.

- To keep up with technological, scientific, economic, and financial developments.
- To transfer knowledge and technology to the Arab world.
- To modernize Arabic by adding new concepts and terms through coinage and arabization.
- To make information retrieval, extraction, and translation easier for the Arab user to bridge the gap between people in the Arab world and their peers in more technically advanced countries.

According to White (as cited in Zantout & Guessoum, 2000), "a country unable to assimilate a high volume of potentially useful information from abroad could lack timely, accurate data and lose its edge in international business, diplomacy, military readiness, and academic research" (p. 123).

Challenges for Arabic MT

Arabic is a very rich language that belongs to a different language family, namely the Semitic languages, than the languages spoken in the West, which are mainly Indo-European. So, when translating between Arabic and Western languages using MT, many problems arise that make it difficult for machines to produce good outputs.

Habash (2007) states that "Arabic has a very rich morphology characterized by a combination of templatic and affixational morphemes, complex morphological rules, and a rich feature system" (p. 263). According to Farghaly and Shaalan (2009), these features are as follows:

- Arabic is written from right to left.
- There is no capitalization in Arabic, which makes it difficult to recognize names of entities.
- There are no rules for punctuation. Actually, there are rules, but there is lack of training.
- It has a flexible word order.
- Its letters change shape according to their position in the word.

- Some Arabic letters share the same shape and are only differentiated by adding certain marks such as a dot, a *hamza*, or a *madda* placed above or below the letter.
- It allows subject pronouns to be dropped.
- It has a very rich and complex agreement system. A noun and its modifiers have to agree in number, gender, case, and definiteness.

Izwaini (2006, pp. 121-129) notes the following features:

- Many Arabic names have meanings.
- Arabic sentences can be nominal (subject—verb) or verbal (verb—subject), whereas English sentences are mainly nominal (subject—verb).
- Arabic uses constructions that literally mean *friend of*, *mother of*, or *father of* to indicate ownership, a characteristic, or an attribute.
- Arabic uses pronouns of two genders only; it has no gender-neutral pronouns.
- The absence of diacritics "*al tashkiil*" can change the meaning.
- In Arabic, possessive pronouns (one or two letters) are attached to nouns.
- There are no copula verbs in Arabic.
- In Arabic, generic names, many place names, and titles have to be used with *al*, which functions as a definite article.

These features present many ambiguities for machines. Moreover, other important and more sophisticated problems must be addressed. Therefore, more refined rules and possible solutions need to be established for MT than are currently available so that it can handle Arabic and produce more accurate translations.

A good MT system should endeavor to go a step beyond the gist level. Rules need to be developed and refined so that the output can match the finest possible product with only a minimum of post-editing. The less post-editing is required, the more successful the translation will be and the less time and work will have to be invested to produce the final translation (Izwaini, 2006, pp. 146-147).

Arabic MT Systems

Sakhr systems. Sakhr is the first Arabic company to deal with Arabic language technology in more than 20 years. It provides a bidirectional Arabic-English MT system, which is a hybrid engine that optimizes rule-based and statistically based processes to achieve rapid and accurate translation (Shakhr Software, n.d.a). Sakhr is the only company in the Arab world with an interest in MT; it has conducted a good deal of research in this field and built some systems that translate from and into Arabic—all other systems that do this kind of work are built in the West (Zughoul & Abu-Ashaar, 2005, p. 1031).

According to the Sakhr Software (n.d.b) Website, Sakhr can also be customized for specific domains thanks to decades of natural language processing (NLP) research on the Arabic language. It offers the following features: online translation, speech-to-speech translation, mobile translation, news and media monitoring, professional translation services, and many more.

According to Zughoul and Abu-Ashaar (2005), "apart from one company [Sakhr] in the Arab world, there has been no serious work on the development of software packages of MT or other applications of NLP for Arabic in its areas of influence, i.e., where it is a first language" (p. 1031). Following is an outline of Sakhr MT systems:

- **Sakhr Enterprise Translation (SET) V. 5.1:** A Web-based MT system that translates from Arabic into English and vice versa. SET V. 5.1 is a transfer-based MT engine that performs extensive analysis on morphological, lexical, syntactic, and semantic levels with the aim to understand the input text and, hence, resolve all kinds of ambiguities prior to translation. SET V. 5.1 is designed for companies and organizations that have many translation tasks, and it also translates documents and Websites. A dictionary of medical terms was added to this version (Aramedia, n.d.).
- **Tarjim:** A free online MT service that translates from Arabic into English and vice-versa. Tarjim (n.d.) translates texts and Web pages. It is also mentioned on the Ajeeb (n.d.) Website together with other services, including Arabic, English, and French

dictionaries, six glossaries, and a spelling checker. The Tarjim tool can be downloaded free of charge, but a monthly payment is required so that it can be used.

ATA software systems. ATA is a London-based software house, specializing in Arabic business software. It was established in 1992 by programmers and Arabic software specialists (ATA Software, n.d.a).

- **Al Mutarjim Al Arabey V. 3.0:** A bidirectional English-Arabic MT system that translates a variety of texts and is the latest version developed by the ATA Software Company for professional users. Al Mutarjim Al Arabey V. 3.0 can be tuned in many ways to enhance and transform the translation quality. It has 46 specialized science dictionaries in medicine, veterinary, science, biology, physics, mathematics, chemistry, engineering, and geology (see Figure 5). It has a dictionary of abbreviations, and it can transliterate proper names. Users can add their own dictionaries. Al Mutarjim Al Arabey V. 3.0 can translate more than 10,000 words per minute (ATA Software, n.d.d).

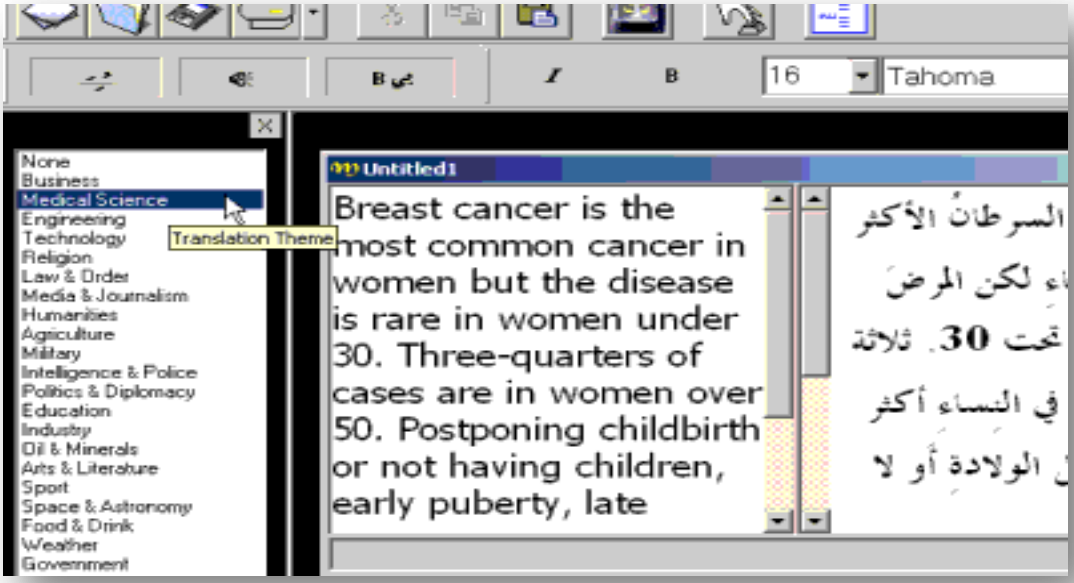

Figure 5. Al Mutarjim Al Arabey—translation theme. [screenshot]. From *Almutarjim Al Arabey, V. 3.0*, n.d.d. by ATA Software. Copyright 1995 by ATA Software Technology, Ltd. Reprinted with special permission.

- **Al Wafi V. 4.0:** A bidirectional English-Arabic MT system (see Figure 6) designed for nonprofessional users. Al Wafi V. 4.0 is easy to use and ideal for people who need fast translations. It contains an English dictionary with more than 200,000 entries and an Arabic dictionary with more than 75,000 entries, and the user can create his or her own dictionary. It transliterates proper names and translates up to 1,000 words per minute. It has an English spelling checker (see Figure 7) and an optional voice output (see Figure 8; ATA Software, n.d.b).

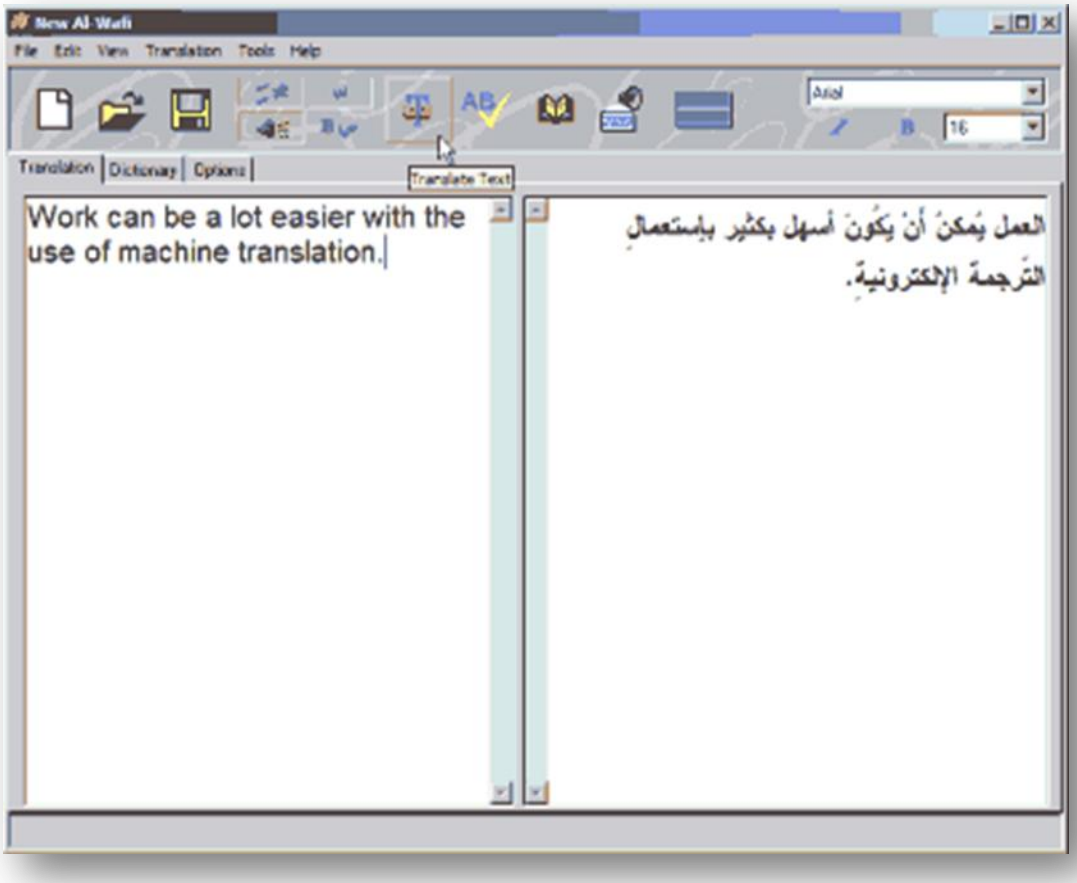

Figure 6. Al Wafi translation. [screenshot]. From *Al Wafi, V. 4.0*, n.d.b, by ATA Software. Copyright 1995 by ATA Software Technology, Ltd. Reprinted with special permission.

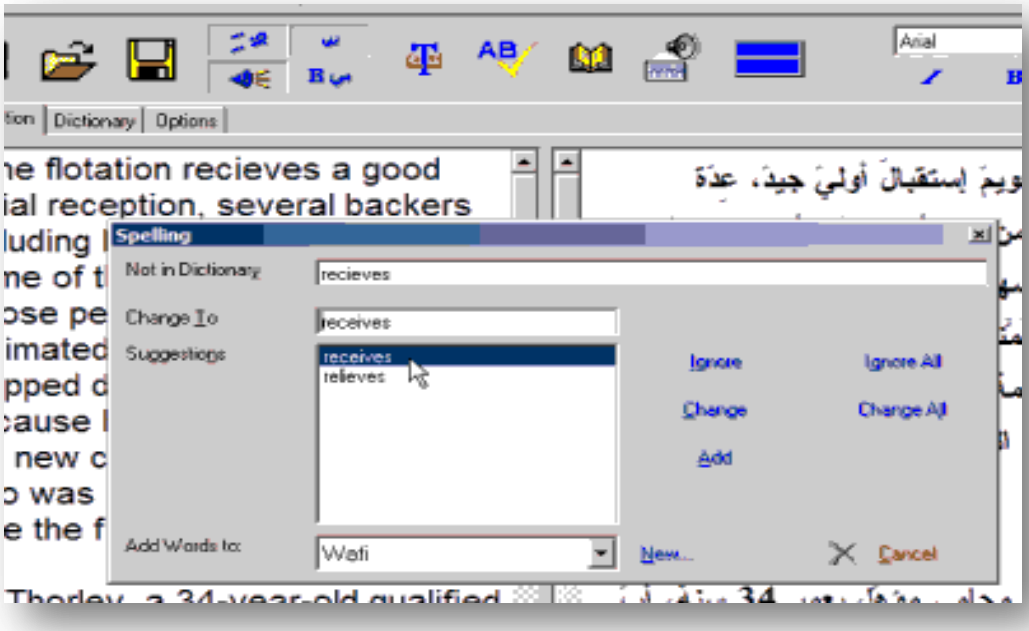

Figure 7. Al Wafi—spelling checker. [screenshot]. From *Al Wafi V. 4.0*, n.d.b, by ATA Software. Copyright 1995 by ATA Software Technology, Ltd. Reprinted with special permission.

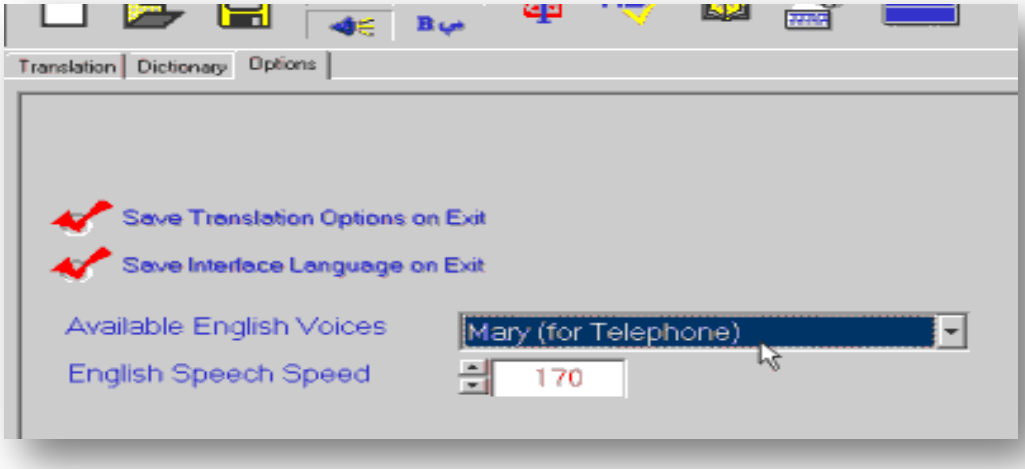

Figure 8. Al Wafi—voice output. [screenshot]. From *Al Wafi, V. 4.0*, n.d.b, by ATA Software. Copyright 1995 by ATA Software Technology, Ltd. Reprinted with special permission.

- **Golden Al Wafi V. 2.0:** A bidirectional English-Arabic MT software for advanced users and professional translators. Its expanded and specialized dictionaries make it suitable for more advanced translation. Its dictionary contains more than 2 million English and Arabic entries. It has eight specialized science dictionaries in medicine, veterinary science, biology, physics, mathematics, chemistry, engineering, and geology (see Figure 9), and the user can create his or her own dictionary. It transliterates proper names and translates up to 1,000 words per minute. It has an English spelling checker and an optional voice output (see Figure 10). It can perform multidocument translations (see Figure 11). It can also display Arabic texts with diacritics. Version 3.0 was launched online in 2007. Access is granted via monthly or yearly subscription (ATA Software, n.d.e).

Figure 9. Golden Al Wafi—specialized dictionaries. [screenshot]. From *Al Wafi V. 4.0*, n.d.e, by ATA Software. Copyright 1997 by ATA Software Technology, Ltd. Reprinted with special permission.

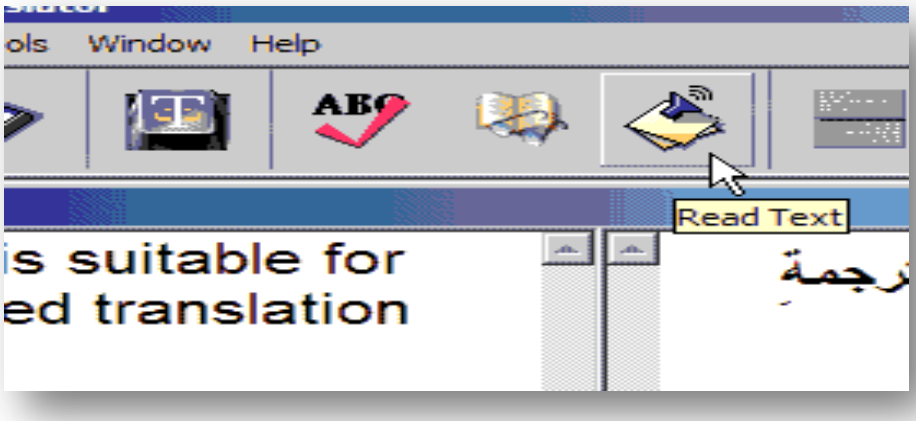

Figure 10. Golden Al Wafi—voice output. [screenshot]. From *Al Wafi V. 4.0*, n.d.e, by ATA Software. Copyright 1997 by ATA Software Technology, Ltd. Reprinted with special permission.

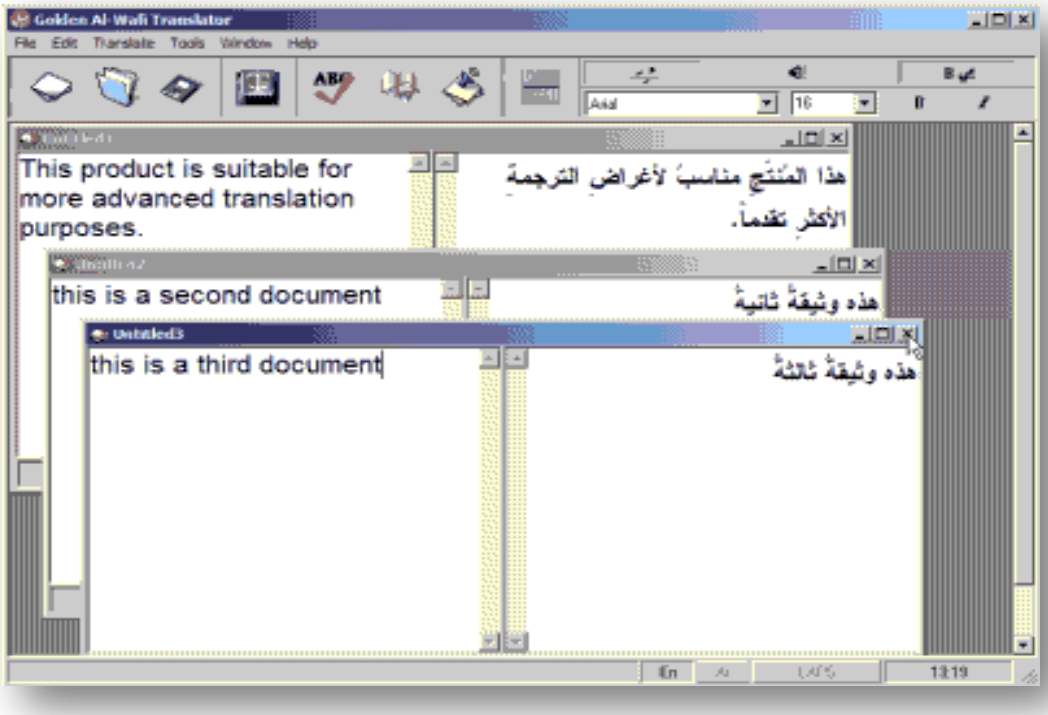

Figure 11. Golden Al Wafi—multidocument translation. [screenshot]. From *Al Wafi V. 4.0*, n.d.e, by ATA Software. Copyright 1997 by ATA Software Technology, Ltd. Reprinted with special permission.

- **MutarjimNet V. 1.0:** Designed for companies, it provides a translation network for their employees. Its dictionary contains more than 2.5 million entries. It has over 40 specialized dictionaries, which are look-up only. It translates highlighted text from the Internet and from Word and PDF documents as well as e-mails (see Figure 12). It translates up to 1,000 words per minute (ATA Software, n.d.f).

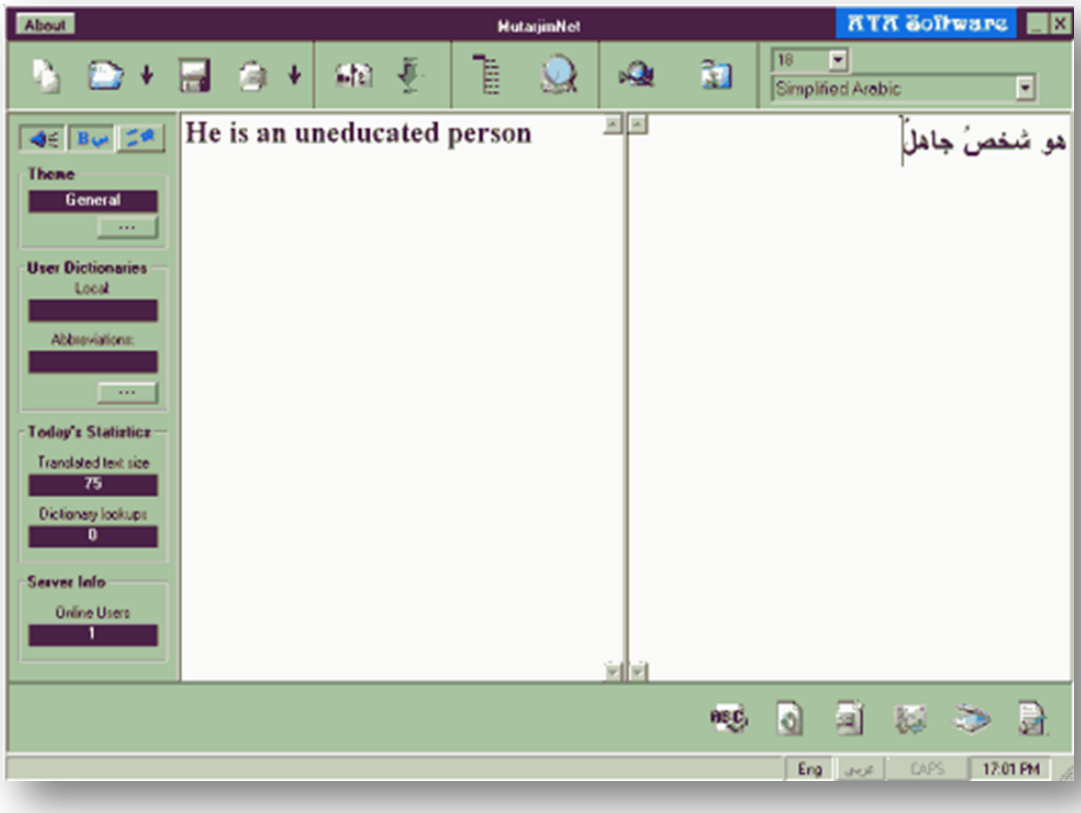

Figure 12. MutarjimNet. [screenshot]. From *MutarjimNet V. 3.0*, n.d.f, by ATA Software. Copyright 1995 by ATA Software Technology, Ltd. Reprinted with special permission.

- **Al Misbar:** An English-Arabic unidirectional online MT system that provides free and instant translation of texts and Web pages (ATA Software, n.d.c; Almisbar, n.d.). Al Misbar has replaced MutarjimNet in 2012.

Cimos systems. Cimos (n.d.b) is an international language service, located in Paris. It pursues two lines of business: conventional translation services and design of multilingual software.

- **Al Nakel Al Arabi:** A bidirectional Arabic-English MT system with specialized dictionaries in computing, banking, business, gas and petroleum, medicine, finance, and aerospace, among others. It makes use of a rule-based system and a knowledge database. Al Nakel Al Arabi (Cimos, n.d.a) has a parser; it performs deep syntactic analysis and selective semantic analysis to detect main verbs, phrasal verbs, and idioms. Its dictionary contains 150,000 words and phrases. It translates up to 60,000 words per hour. It is available as a stand-alone product or a Web-based server (Cimos, n.d.a).
- **Multilingual Translation System (MLTS):** A Web-based MT system that can serve both as a tool to assist in the human translation process and as automated translation software to provide a rough draft translation. It translates about 100,000 words per hour (Cimos, n.d.c).

AppTek systems. Applications Technology Inc. (AppTek) is a U.S. company that has been specializing in translation software for 20 years. (AppTek, n.d.a). AppTek is now part of Science Application International Corporation (SAIC, 2010).

- **TranSphere:** A hybridized MT system of two fully integrated engines that use statistics-based and rule-based MT (see Figure 13). It runs on stand-alone work stations, client-server architecture, or as a Web-based translation system of online content. This system integrates hardware and software that provide
  - rule-based MT,
  - statistics-based MT, and
  - hybrid MT.

TranSphere (AppTek, n.d.b) supports bidirectional translation from and into different languages, including Arabic; it translates documents and Web pages. It has a general dictionary that contains more than 100,000 entries. It also uses special microdomain dictionaries as, for example, law, business, accounting, finance, military,

computing, electronics, petroleum, economics, and telecommunication, among others. It can be used by

- a multilingual corporation with worldwide manufacturing operations to save time and cost when translating, for example, manuals for facilities in different countries;
- professional translators, who can use it to assist them, save time, and increase their productivity; and
- government agencies that deal with large amounts of documents that need to be translated (AppTek, n.d.b).

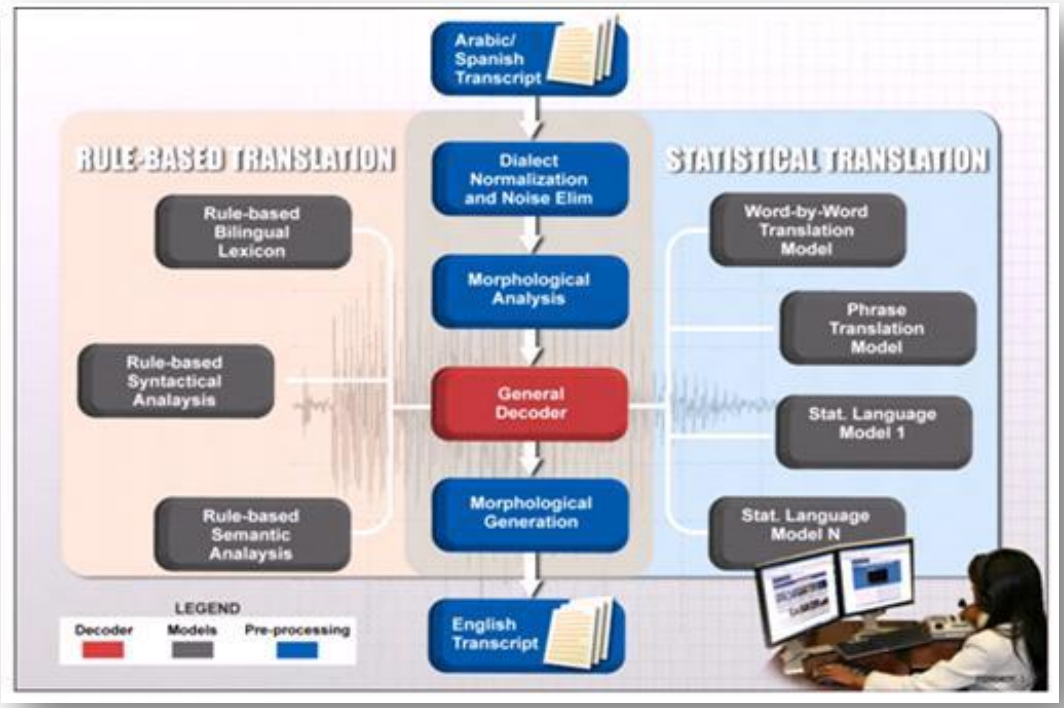

Figure 13. AppTek MT system TranSphere. From *TransPhere machine translation system*, n.d.b, by AppTek. Copyright 2003 by AppTek/SAIC. Reprinted with special permission.

- **WebTrans:** WebTrans is a bidirectional Web-based MT system that is based on AppTek's (n.d.c). TranSphere MT technology. It translates Web pages and allows online access to Web content while maintaining Website format integrity (see Figure 14). It enables

Website localization, online news service monitoring, and access to MT anywhere a Web browser is available. It has access to special domain microdictionaries such as those created for computers, aviation, the military, telecommunications, economics, diplomacy, nuclear uses, drugs, chemicals, law enforcement, and business (AppTek, n.d.c).

Figure 14. Sample of MT of a newspaper front page by AppTek WebTrans. From *WebTrans machine translation system*, n.d.c, by AppTek. Copyright 2003 by AppTek/SAIC. Reprinted with special permission.

SYSTRAN. SYSTRAN is located in Paris and, according to its Website, has been a leading supplier of language translation software for more than 40 years. It translates from and into 52 languages for individuals and companies. In 2009, it introduced the first hybrid MT engine, which combines the advantages of technology used in linguistics with statistical techniques; thus, the software automatically learns from existing and validated translations. The self-learning technique allows users to train the software to any specific domain to produce translations.

SYSTRAN is used by global corporations such as Symantec; Cisco; EADS; and Internet portals such as Yahoo, Lycos, and AltaVista, as well as public agencies such as the U.S. intelligence community and the European Commission (SYSTRAN, n.d.a). SYSTRAN offers the following systems:

- **SYSTRAN 7 Home Translator:** This allows the user instantly to translate and understand Web pages, e-mails, instant messages, conversation forums, letters, documents, and other text (SYSTRAN n.d.d);
- **SYSTRAN 7 Office Translator:** Offers integration with Microsoft Office. It also automatically learns new terminology to improve the quality of its translations (SYSTRAN , n.d.e);
- **SYSTRAN 7 Business Translator:** Creates, understands, and manages multilingual information for employees, customers, prospects, partners, and vendors worldwide. It has a specialized business dictionary (SYSTRAN, n.d.b);
- **SYSTRAN 7 Premium Translator:** Designed to meet the needs of professionals and corporate users; it enables the user to produce translations for multilingual documents. It also allows the user to add specialized dictionaries (SYSTRAN, n.d.f);
- **SYSTRAN Enterprise Server 7:** A comprehensive solution able to meet a wide range of translation needs on an enterprise scale. It is available in three editions: workshop, standard, and global (SYSTRAN, n.d.c);

SYSTRAN also offers individuals and businesses translation services of Web pages, texts, or Microsoft Office files:

- **SYSTRANNet:** Free of charge, it provides real-time translation of texts, Web pages, and personal files (SYSTRAN, n.d.i);
- **SYSTRANBox:** Can be added to personal Websites or portals in order to communicate in multiple languages and understand foreign language content in real-time by translating personal or business texts or any information (SYSTRAN, n.d.g); and

- **SYSTRANLink:** Localizes Websites and content applications such as search tools in real time. Visitors to Websites can choose their native language and see the information in that language within a short time (SYSTRAN, n.d.h).

All SYSTRAN (n.d.a) products use the same translation engine that is used by Internet portals, global corporations, and government agencies, and all are available to translate from and into Arabic.

Other Web-Based MT systems. ***Babylon*****.** Babylon (n.d.a) is a computer dictionary and translation program, developed by Babylon Ltd. It includes built-in dictionaries. It has developed 35 English-based proprietary dictionaries in 19 languages, which have between 60,000 to 200,000 terms. It includes monolingual glossaries in over 75 languages, which are indexed in 400 categories covering the arts, business, computers, health, law, entertainment, sports, and more. It also has voice output. It provides full text, full Web page, and full document translations in 33 languages, including Arabic. It supports integration with Microsoft Word. It offers results from a database of over 1,600 sources in 75 languages. In 2011, it launched an online question-and-answer translation software (Babylon, n.d.a; Babylon, n.d.b).

Google Translate. Google Translate is a free statistical online MT service with voice output, provided by Google Inc. to translate texts, documents, and Websites from and into 58 different languages, including Arabic. It allows the user to provide suggestions and corrections to the translation. It also provides a Translator Toolkit to assist translators into which the user can upload his or her translation memory (Google Translate, n.d.a; Google Translate, n.d.b).

Bing Translator. Bing Translator is a free online statistical MT service provided by Microsoft. It translates texts and Web pages from and into 37 different languages, including Arabic. It allows the user to provide suggestions and corrections to the translation. It provides additional functionality such as detection of the language of a given text, but it does not have voice output (Bing Translator, n.d.a; Bing Translator, n.d.b).

TRADOS

Translation and Documentation Software (TRADOS) is a suite of tools that includes mainly a translation memory (TM). TM is defined by Trujillo (1999) as a translation aid that searches a database of previously translated text fragments, known as translation units (TUs), in order to find a similar fragment to the one being translated. Once found, its translation is offered as a draft translation that can be accepted as is, modified, or rejected by the translator. The newly translated fragment can then be added to the database. Initially, the database is empty, but as more texts are translated, the database grows.

A database may also be built automatically from existing translations together with their original texts (Trujillo, 1999, p. 60). As in TRADOS, TM is often integrated into suites of tools that contain, for example, auxiliary programs, a terminology manager, and interfaces to MT systems (Kenny, 2011, p. 469). TRADOS currently supports Google Translate and Microsoft Translator. Although MT systems can be used only for a limited number of supported languages, TMs such as TRADOS can be used for an unlimited number of supported languages. TRADOS is also compatible with a wide range of programs used to create content, including Microsoft Office, OpenOffice, RTF, Tab Delimited, HTML, and XML (SDL TRADOS, n.d.). Moreover, because not many exact matches for the fragments can be expected to be found in the TM, TRADOS employs a feature known as *fuzzy matching*, which is able to find fragments in the database that are an approximate or partial match for the fragments in the new source text. TRADOS can also be networked so that multiple translators can search and contribute to the same database (Bowker & Barlow, 2008, pp. 2-3).

TM is seen as a productivity-enhancing resources; so, if a sentence is already translated, there is no need to translate it again and the translation in the database can be reused. It is also seen as a quality-enhancing resource in translations where consistency is at a premium (Kenny, 2011, p. 464). Moreover, TM is preferred by translators because humans have complete control over the translation process.

Despite the many benefits of TM, it also has many drawbacks. If the contents of TM are of poor quality, then the new translations will be of poor quality as well, unless translators are willing to spend more time correcting the translations. TM works by matching text fragments; therefore, translators translate sentence by sentence rather than by changing the order of sentences or by combining or splitting sentences, which may lead to the production of less cohesive target texts. The new TM is useless because it is empty, and it takes time to build the database. Moreover, upgrading the hardware or installing a network to allow multiple translators use the same TM can be very expensive. TRADOS, which is very popular, is well-known for its complexity, and it may take a translator several weeks to feel comfortable working with it. Some translators find that editing a fuzzy match may take longer than translating a fragment from scratch (Bowker, 2002, pp. 114-118).

Summary

In this chapter, the features of the Arabic language were discussed, and problems for MT systems due to these features were exposed. The chapter also outlined systems that can translate from and into Arabic. To find out how much interest is shown in this kind of translation technology in the Arab world and how much research is being conducted on Arabic MT, a survey was undertaken. In the next chapter, the results of the survey will be discussed and analyzed.

Chapter 4: Analysis and Findings

To examine how widely MT is used in Saudi organizations and how much Saudi universities are interested in teaching and researching MT, an extensive survey was conducted. This chapter discusses the survey and analyzes its findings.

Data

To map the status of MT in Saudi Arabia, data had to be collected from actual and potential users of MT such as organizations and translation agencies as a well as institutions of higher education. Following is the procedure adopted to collect the data:

1. A questionnaire (see Appendix A) was designed to obtain answers to the research questions posed for the study and sent out to 100 Saudi governmental and nongovernmental organizations (see Appendix B), including 12 ministries, 27 governmental organizations, 50 of the biggest private companies and nongovernmental organizations in the country, four banks, four hospitals, and three newspapers.

2. To investigate the status of MT in the business affairs of translation agencies, what these agencies think about MT, and whether they use any of the available systems, data were collected from 30 translation agencies, including translation centers at both King Saud University (KSU) and King Faisal University (KFU; see Appendix C).

3. To find out how much Saudi universities are interested in MT, a questionnaire (Appendix D) was specifically designed for higher education institutions to obtain answers to the research questions and sent to all 21 universities that offer degrees in languages (Appendix E). More data were collected from the university Websites. Data also were collected from the Websites of 28 universities that offer degrees in Computer Science and Information Technology (Appendix F) about faculty who are interested in MT and related fields such as natural language processing (NLP) and Arabic computing, as well as on research conducted in these fields.

A number of important limitations on the process of data collection must be considered. First, the number of responses from organizations, translation agencies, and universities was relatively small and not as abundant as had been hoped (this will be further elaborated in the next section). Follow-up contact was, therefore, required for

each of these entities in order to increase the response rate. This also proved to be rather difficult because many of them neither replied to e-mail nor to phone calls. Second, it was not possible to conduct personal interviews as women are not allowed to visit governmental organizations, unless there is a women's section. However, this is not very helpful because such sections are usually small and entrusted with limited duties. Third, Saudi Arabia is a big country with many organizations and companies, and it is very difficult, if not impossible, to cover all of them within the scope and the time allotted to a study of this kind, conducted under the auspices of a university. The study was, therefore, limited to 100 organizations and companies. Lastly, it was difficult to obtain data on research conducted on MT because some faculty members and university librarians did not respond or because university Websites lacked pertinent information.

Analysis

The findings of the survey are analyzed in three sections: MT in Saudi organizations, MT in Saudi translation agencies, and MT in Saudi universities and research centers.

MT in Saudi organizations. A total of 66 responses were received from six ministries, 22 governmental organizations, one newspaper, one hospital, one bank, and 35 private companies. Only 44 responses (66.7%) were analyzed because the other 22 organizations (33.3%) do not need translation (see Figure 15) for one of the following reasons:

- They use only one language, either Arabic or English.
- Their employees are bilingual.
- They outsource their translations to translation agencies.

Out of those 44 organizations, 32 (72.7%) need translation on a continuous basis. Thirty three of the 44 (75%) have either a translation department or a translator, whereas the other 11 organizations (25%) rely on their bilingual employees and sometimes outsource their translations to translation agencies. This is illustrated in Figure 15.

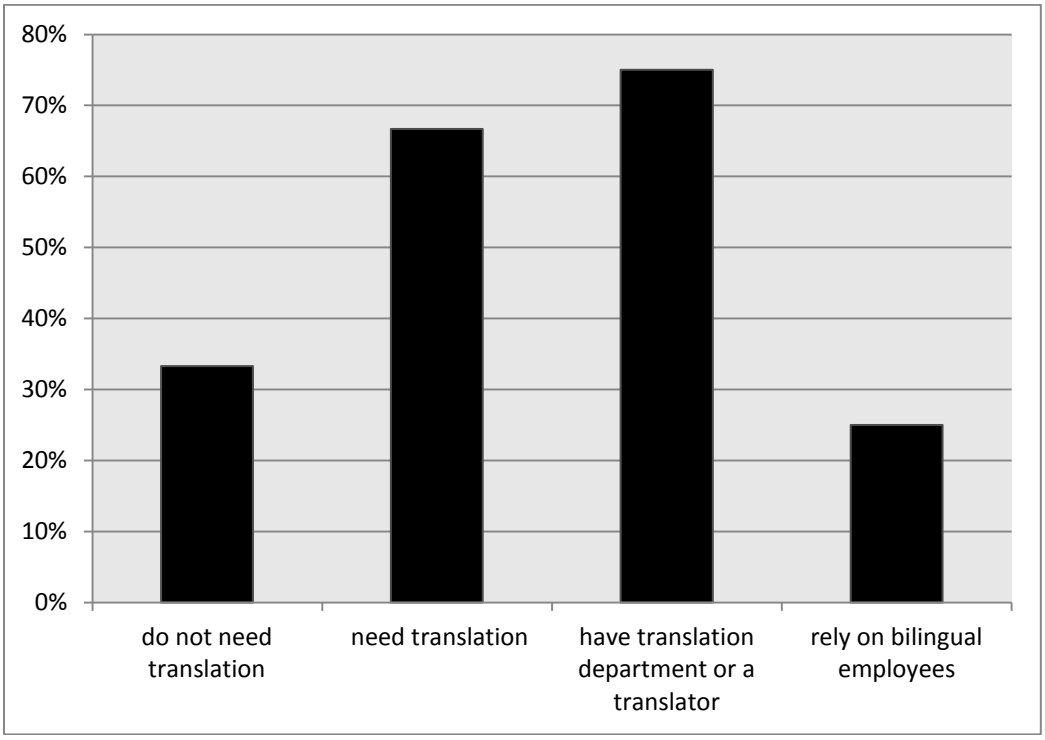

Figure 15. Status and use of translation in Saudi organizations.

Only 17 of the 44 organizations (38.6%) use MT (see Figure 16). These 17 organizations use online translation, specifically Google Translate. Only two organizations use both Google Translate and Babylon translation software. No other translation software has been used by any of the organizations that responded, except by Shura Council, which has previous experience with different translation software over the last 10 years. This will be discussed further in an upcoming section.

Eleven of the aforementioned 17 organizations (64.7%) use MT to translate different types of texts, including official documents. Twelve (70.6%) conduct post-editing of the translation output, whereas the remaining five organizations (29.4%) do not because they need only a gist of the content or no final versions of translated documents are needed.

Twenty-seven of the 44 organizations (61.4%) do not use MT (see Figure 16) for one of the following reasons:

- They do not trust MT (63% out of 27).

- They cannot afford it (3.7% out of 27).
- Their texts are in hard copy (3.7% out of 27).
- They believe that MT output is inaccurate (11.1% out of 27).
- They think that, with post-editing, it makes for twice the work (3.7% out of 27).
- They do not need it because they either do little translating or rely on their own translation departments or translators (3.7% out of 27).

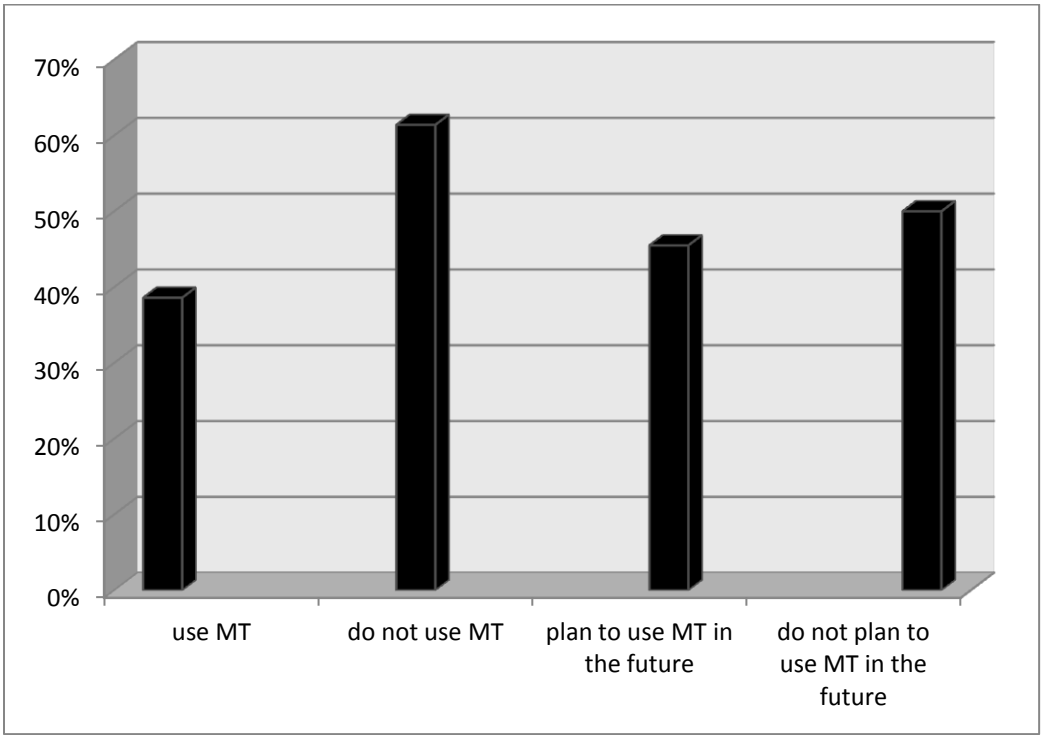

Figure 16. MT in Saudi organizations.

Twenty of the 44 organizations (45.5%) plan to continue using MT or to start using it in the future (see Figure 16). Seven of these 20 organizations (35%) provided one or more of the following comments:

- Google Translate is really good for the translation of words.
- Google Translate is much easier and faster than regular dictionaries.

- They plan to use MT software during a trial period.
- MT produces only draft translations. That means, it is only 50% of the translation process. The output needs much reviewing and editing, but it does save time[1].

Twenty-two of the 44 organizations (50%) do not plan to continue using MT or to start using it in the future (see Figure 16). Nine of these 22 organizations (40.1%) gave one or more of the following reasons:

- MT is not reliable and TM (such as TRADOS) is a better alternative; however, it requires a large financial investment.
- Google Translate is good but only for words. MT may translate simple sentences but not complex ones.
- MT cannot be used for translating technical and legal documents.
- MT is not accurate; so it cannot be used to translate official documents.
- MT produces literal translation and, therefore, provides strange output.
- MT cannot deal with abbreviations or understand culture-specific aspects.
- MT can be used only to get a general idea about a text. Translation is not science; it is an artistic task that can only be done by humans.
- MT software cannot be considered really viable, unless a significant breakthrough is achieved in terms of technology. A computer can be easily fed with all kinds of different styles, jargon, cultural voids. It is doable, but what about implicatures?[2]
- Arabic differs in terms of its morphology, syntax, and semantics from Western languages, which makes it difficult for machines to translate from and into Arabic (as discussed in chapter 3).
- MT needs post-editing, which means double-work.

Two of the 44 organizations (4.5%) have no plans for future MT use and are waiting to see how much improvement will be achieved.

Shura Council experience with MT systems. Because the Shura Council is the only organization taking part in the survey that has used different MT systems, it was of

[1] Shura Council response
[2] Aramco response

special interest to consider what its staff had to say about these systems and to what degree they found them helpful.

The Shura Council has a translation department with six staff members who are constantly engaged in translating official documents, reports, paperwork, and e-mail. Ten years ago, they started using MT with the English-into-Arabic-only Golden Al Wafi (discussed in chapter 3) whose translation quality, according to Shura staff, ranged from 10% to 40%. Then, MutarjimNet (also discussed in chapter 3) was offered to the Council by the manufacturing company that claimed that its vocabulary and database of terms were connected to a UN database. It turned out that either the claim was untrue or the system was useless. At the same time, other software was used such as Ajeeb (or Tarjim, described in chapter 3) and Easy Lingo (English-Arabic dictionary software), both of which proved to be unhelpful, especially when dealing with names of local and international organizations. So, the Council ended up with Golden Al Wafi (ATA Software, n.d.e), which proved, at the time, to be better than the others.

According to Shura Council staff, Google Translate (described in chapter 3) was and still is much better than the others, especially, when dealing with names of local and international organizations, international treaties, specialized economic terms, and even names of herbs and diseases. They also noted that Google Translate seems to be developing rapidly, in contrast to other software.

In recent years, more improved software has appeared on the market and was used by the staff, for example, Sakhr Enterprise Translation (SET; explained in greater detail in chapter 3), which has seen more improvements than MutarjimNet. The former offers more tools to assist the translator, and the translation quality is more desirable.

At this writing, Shura Council staff use only Google Translate with post-editing of the output, and, according to them, Google Translate has proven to be the best available system. It is continually being developed, even day by day. Its translation accuracy may range between 40% and a high of 90%, but it usually ranges between 70% and 80%.

Discussion. Based on the responses received from the 44 organizations that took part in the survey, it became clear that 61.4% of them do not trust MT. Most of them have already decided that MT is not accurate, or not trustworthy, even those who have never tried it (based on their responses to the questionnaire). In addition, some hold views that are clearly incorrect, for example, MT cannot deal with abbreviations and technical and legal texts. As a matter of fact, abbreviation dictionaries are included in MT software (and can even be added to), and technical texts are one text type that is well-suited to being translated by MT. As for legal texts—although the point was made in chapter 2 that this text type is difficult for MT, at least according to Melby (1987, p. 146), due to its complex syntax and need for accuracy—they, in fact, display language that is quite formulaic and subject to rigid rules of text arrangement; thus being suitable for MT. Some translators find post-editing a tiresome job, although they are supposed to review any translation whether done by machine or by themselves. Quah (2006) notes that, "machine translation is an important technology which has social, political, commercial and scientific benefits though there are many misconceptions about its success or failure" (p. 57).

MT can be very helpful for organizations. It translates fast and keeps translation consistent. It covers all domains, is confidential and available any time. MT is not expected to be used to translate novels, poems, or advertisements. Yet, it can produce a rough translation that can be used to get a general idea about a text, for assimilation, interchange, or database accesses (as described in chapter 2). It can also be used to translate repeated simple sentences and text types that do not include complicated structures or culture-specific expressions such as manuals, weather forecasts, reports, and more. According to the SYSTRAN Website, MT may not be able to produce good translations of literary or cultural texts, but it can definitely help translators in translating the tedious and repetitive parts in scientific and technical documents, commercial and business transactions, administrative memoranda, legal documentation, instruction manuals, agricultural and medical text books, industrial patents, publicity leaflets, and newspaper reports.

Only 11.4% of the 44 organizations that responded to the survey need translation of official documents, meaning, they require accurate translations, not merely rough drafts. The other 88.6% need translation of different text types such as reports, general paperwork, e-mail, and Websites, for which a rough translation is sufficient.

As for Arabic, it is true that machines can face difficulties when translating between Arabic and Western languages due to the complicated morphology, syntax, and semantics of Arabic. However, if more refined rules for Arabic morphology and syntax are integrated in MT systems, those difficulties, or at least most of them, can be overcome. Farghaly and Shaalan (2009) observe that "Arabic NLP applications must deal with several complex problems pertinent to the nature and structure of the Arabic language" (p. 2).

Most notable in the responses of 76.4% of the organizations that use MT is that they use Google Translate as an online dictionary or to translate simple sentences. Google Translate is popular because it is free, easy to use, and fast. It translates between 58 languages. It is online software that is easy to access and more practical than others when translating Websites and e-mail. Its translation quality can be improved by using its Translator Toolkit or by uploading the user's own TM into the Translator Toolkit.

MT in Saudi translation agencies. Fifteen out of the 30 translation agencies (50%) included in the survey responded. Only six of these 15 agencies (40%) use the online MT Google Translate and only to translate words or simple sentences (see Figure 17). They found that it is fast; so it saves time, especially with typing. According to these agencies, Google Translate is very accurate when translating abbreviations and the names of local and international organizations, and more. However, they found it to be 90% inaccurate when translating complex text and capable of being misleading. Getting a rough translation can be useful; however, depending on the text type, only experienced translators can get full benefit from using it.

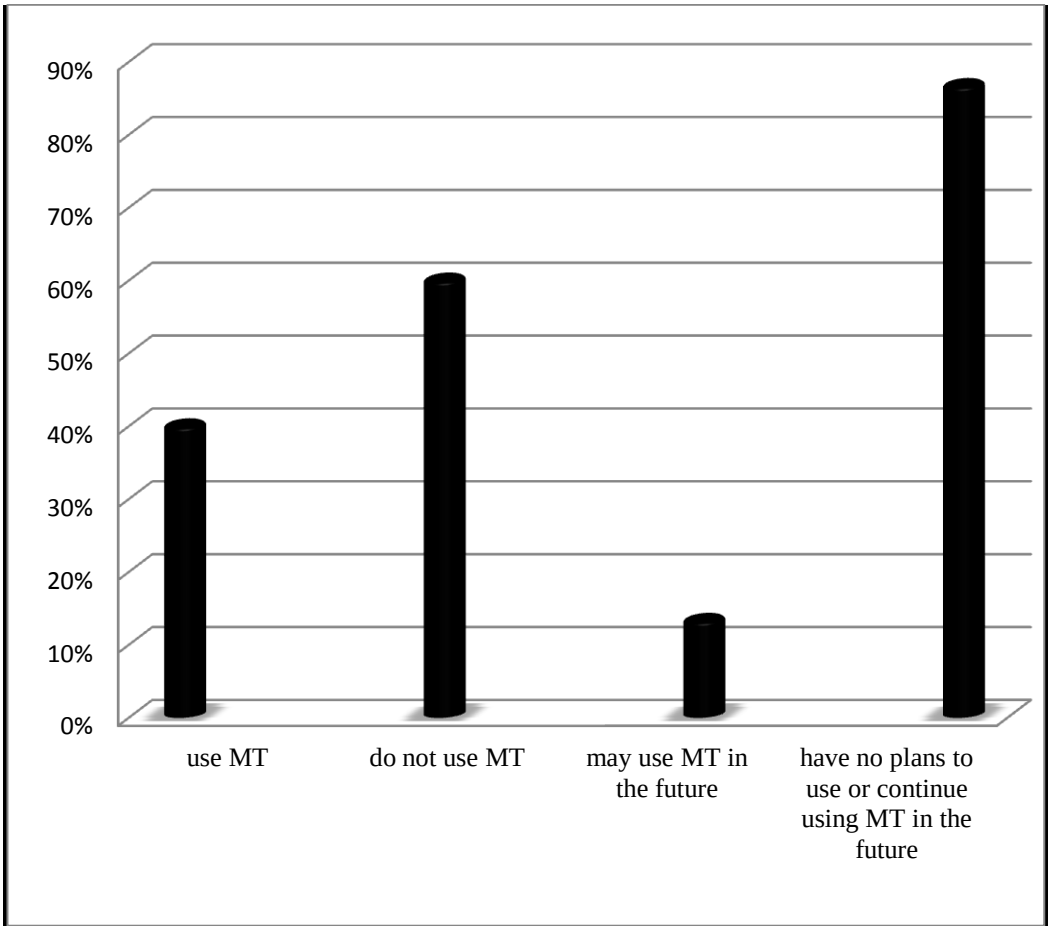

Figure 17. MT in Saudi translation agencies.

The King Faisal University's (KFU) translation center and the translation center of the Riyadh Chamber of Commerce and Industry also use Babylon software but prefer Google Translate. The King Saud University's (KSU) translation center does not use MT because there is not enough interest expressed by staff and faculty to try this technology.

Nine of the 15 agencies (60%) completely reject the idea of using MT (see Figure 17) because, as they maintained, it is not advanced enough; it produces a poor quality of translation; and it needs post-editing, which means double the work. Some even think that MT is a joke.

Thirteen of the 15 translation agencies that responded to the survey (86.7%) have no plans to use or continue using MT. Two of the 15 agencies (13.3%) will consider using it in the future (see Figure 17) under one of the following conditions:

- If their business expands and they need many translations done.
- If MT systems become more advanced.

Discussion. It is not surprising that translation agencies do not use MT, or when they do decide to use it, that they opt for Google Translate to translate words only because they believe that they have to produce accurate translations. Their job is not to give people a general idea about a text or a rough draft. Any bilingual person can do that, but not every bilingual person can provide an accurate translation of a given text. Only a professional translator can do that. However, translators can utilize the help that MT offers and, at least, get rid of all those paper and even electronic dictionaries to settle down with one comprehensive dictionary and save time. Paper and electronic dictionaries will always be there if needed. Of course, MT has many drawbacks: it is unable to deal with culture-specific aspects and idiomatic expressions, it cannot understand context, it cannot resolve ambiguity, it cannot deal with ellipses or understand colloquial phrases (as explained in chapter 2). However, translators as professionals can take advantage of the benefits of MT (as discussed in chapter 2) and deal with its drawbacks.

Using MT, may make some translators feel threatened and fear for their jobs. According to Vasconcellos (1994, p. 113), there is some truth to this, but a distinction should be made between *existing jobs* and *new work*. Translators should not expect to lose their jobs to MT, but they are expected to learn how to deal with this technology. Lawson (2008, pp. 107-108) observes that MT has led many organizations to hire more translators. He also observes that MT opens up new career avenues as translators learn to edit MT drafts, build MT dictionaries, and identify the improvements needed in the software.

It may be easier for translators who work for or in translation departments in companies and governmental organizations to accept the idea of MT and try it because they usually translate different types of texts, not only official documents, for which a

gist or a rough translation can be sufficient. However, translators who work in translation agencies are used to people or organizations coming to them with official documents and asking for an accurate translation; thus, accuracy becomes their primary goal, and it would be difficult for most of them to accept a rough translation even if it can be post-edited.

MT in Saudi universities and research centers. A large number of universities in Saudi Arabia offer degrees in different fields, and much research is conducted by faculty and graduate and undergraduate students. Because this survey aimed at finding out if Saudi universities are interested in MT, it confined its focus to language and translation colleges and computer science and information technology colleges, where MT is expected to be studied and researched. Because the list of researches in this field is rather long, whether undertaken by faculty or graduate students, it cannot be included here (see Bibliography).

Language and translation colleges. Approximately 21 universities offer degrees in languages. Responses were received from only 12 universities, but more data could be gleaned from these 21 universities' Websites to discover that only six of the 21 universities (28.6%) offer degrees in translation. These are King Saud University (KSU), King Khalid University (KKU), Umm AlQura University (UQU), Imam Muhammad Bin Saud University (IMAMU), Princess Nora University (PNU), and EFFAT University. Three of these six universities (50%) offer or have offered a course in MT, as outlined in the following sections.

- **Imam Muhammad Bin Saud University (IMAMU):** IMAMU offers a course in MT every year for graduate students. Although there is no interest in this field among professors, according to the questionnaire responses, this course is offered every year because there is a plan under way to expand. Three master's theses are being written on MT at IMAMU (see Bibliography).

- **King Saud University (KSU):** KSU offers an introductory course in MT every semester to undergraduate students. No MA theses, PhD dissertations, or other

projects are under way by graduate students because the course is elementary, and no master's program in translation is offered at KSU.

- **Umm AlQura University (UQU):** UQU offered a course in MT only once, and no plan is in evidence to offer it again in the future because of lack of interest among students or professors. No plans exist to expand, and there is no funding available.

Only two professors at KSU are interested in MT. The first has a long career in the area of linguistics and terminology and wrote and published 17 conference and journal papers in MT and Arabic computing (see Bibliography). The second (now retired) taught the MT course at KSU for years. He published a book and 10 conference and journal papers in this field (see Bibliography). He also did a survey (unpublished) on market demand for MT, which proved to be in demand (as provided in the questionnaire responses).

Computer science and information technology colleges. Twenty-eight universities offer degrees in Computer Science and Information Technology. According to the data collected from these colleges, many of them are new and no research has as yet been conducted on MT. Some colleges have no interest in MT, NLP, or Arabic computing. Because NLP and Arabic computing are the fields in which research is needed for developing the tools upon which to build MT systems, the research produced in these two fields had to be included in the survey.

One professor each at Taibah University and at Taif University (TU) showed interest in MT or NLP or both, but no research has been conducted as yet in either place. The following universities have shown more interest in MT and have conducted research:

- **King Abdul Aziz University (KAU):** Three professors are interested in MT and NLP. They presented three conference papers in these fields (see Bibliography). Unfortunately, research at KAU is not funded either by KAU itself nor by King Abdul Aziz City for Science and Technology (KACST). Four MS theses are being written on MT (see Bibliography), and many projects are being done on MT and Arabic computing (although unavailable at the university library or Website, this information was communicated by the respondents via e-mail).

- **King Fahd University of Petroleum and Minerals (KFUPM):** A research group of seven professors work on Arabic computing. Their research is funded by KFUPM and KACST. One professor is a pioneer researcher in MT and Arabic computing. He teaches courses in MT to undergraduate and graduate students. He has published a book and 20 journal articles or conference papers (see Bibliography). Another professor works at the computer research institute of KACST and has published 10 journal articles or conference papers on MT and Arabic computing (see Bibliography). Two MS theses are being written on MT at KFUPM (see Bibliography), and many projects are being carried out on Arabic computing (although unavailable at the university library or Website, this information was communicated by the respondents via e-mail).
- **King Saud University (KSU):** KSU is the first university to show an interest in MT in Saudi Arabia as early as the 1970s, but no research was conducted until the 1990s. About 14 studies on Arabic computing were conducted (see Bibliography). These were funded by KSU or KACST. Two MS theses are being written on MT (see Bibliography), and many projects are being done on MT and Arabic computing (although unavailable at the university library or Website, this information was communicated by the respondents via e-mail).
- **Prince Sultan University (PSU):** Two professors are interested in MT at PSU: one is another pioneer in researching MT and Arabic computing. He has published a book chapter and 15 journal articles and conference papers in these fields (see Bibliography). The second professor has published seven research papers on NLP and Arabic computing (see Bibliography). Currently, PSU does not offer any graduate programs and it does not fund any research on MT.
- **University of Tabuk (UT):** One professor has expressed interest and published 10 journal articles and conference papers (see Bibliography). UT does not fund any research and does not offer any graduate programs.

The most surprising result of the survey is that only one language college, namely KSU, has shown interest in MT and was a research site in this field. However, currently

there is no research being conducted at this college, according to questionnaire responses. No collaboration is taking place between its faculty and the faculty of the computer science and information technology college, who show more interest in MT and conduct more research on MT or a related field, which helps to improve or build Arabic MT systems such as NLP and Arabic computing. Only seven computer science and information technology colleges (25%) have faculty who are interested in MT, and only five out of the 28 colleges (17.9%) conduct research on MT (see Figure 18).

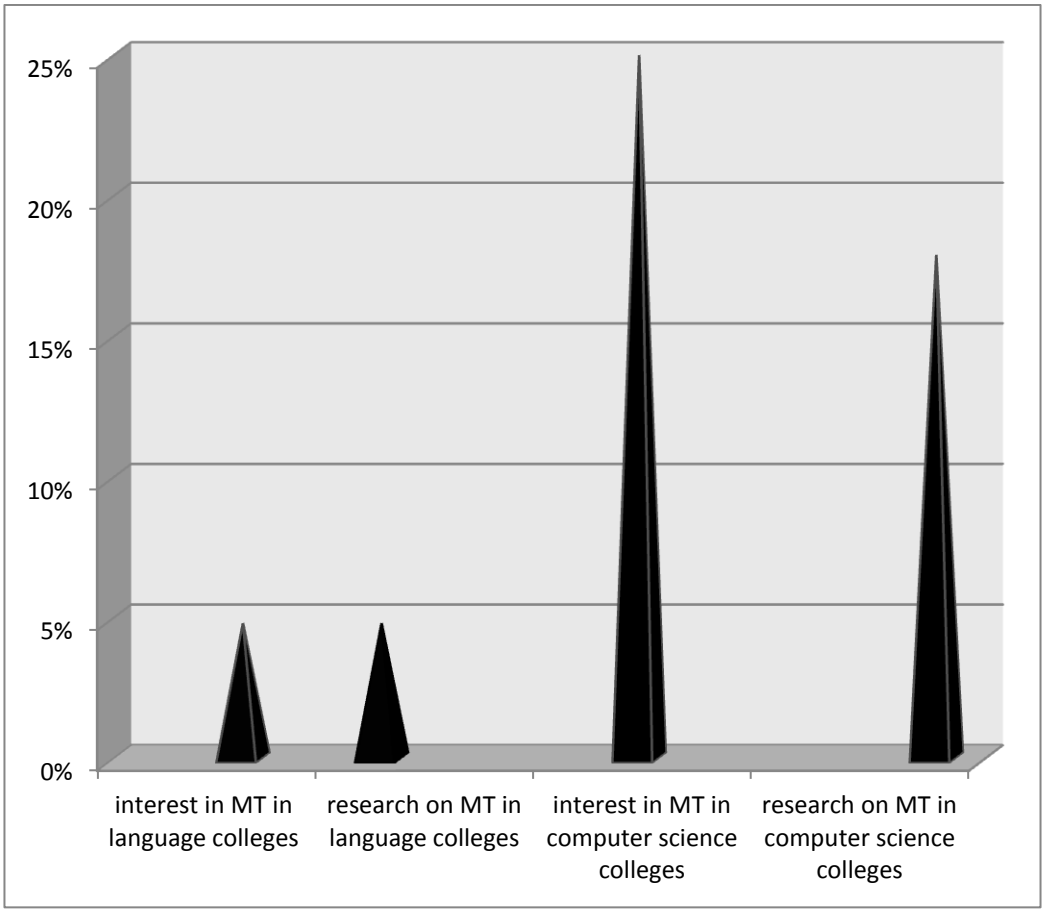

Figure 18. MT in language colleges and computer science colleges.

The survey showed that 109 studies have been conducted. Forty-six of these studies (42.2%) are on MT. Twenty-one out these 46 (45.7%) were conducted in language colleges, and 25 (54.3%) in computer science colleges. The remaining 63 of the 109 studies (57.3%) were conducted on NLP and Arabic computing. Seven of these 63

studies (11.1%) were conducted in language colleges, and 56 (88.9%) in computer science colleges and KACST (see Table 2). Of these 109 studies, 67 (61.5%) are conference presentations, 39 (35.8%) are journal articles, two (1.8%) are books, and one (0.9%) is a book chapter in both language and computer fields (see Table 1).

Table 1

Types and Topics of Studies Conducted on MT, NLP, and Arabic Computing in Language and Computer Science Colleges

Study Type & Topic	Computer Science Colleges		Subtotal	Language Colleges		Subtotal	Total
	MT	NLP & Arabic Computing		MT	NLP & Arabic Computing		
Conference presentations and papers in conference proceedings	15	40	55	5	7	12	67
Journals/ papers	9	15	24	15	-	15	39
Book chapters	1	-	1	-	-	-	1
Books	-	1	1	1	-	1	2
Total	25	56	81	21	7	28	109

The interest in and research on MT in Saudi Arabia started in the 1980s during which time nine studies (8.2%) were conducted, five (55.6%) at language colleges on MT, and four (44.4%) at computer science colleges on NLP and Arabic computing. In the 1990s, the number increased to 37 studies in all fields (33.6%), with 18 (48.6%) conducted at language colleges and 19 (51.4%) at computer science colleges. From 2000

to the time of this writing, the number increased to 63 studies in all fields (57.8%), with five (8%) conducted at language colleges and 59 (92%) at computer science colleges (see Table 2). It should be noted that from the 1980s until this writing, interest and research in MT have decreased in language colleges and increased in computer science colleges.

Table 2

Number of Studies Conducted on MT, NLP, and Arabic Computing Since the 1980s

Field of Study		1980s	1990s	Since 2000	Total
Computer Science Colleges	MT	-	6	19	25
	NLP & Arabic Computing	4	13	39	56
	Subtotal	4	19	58	81
Language Colleges	MT	2	15	4	21
	NLP & Arabic Computing	3	3	1	7
	Subtotal	5	18	5	28
Total		9	37	63	109

Neither interest nor funding is sufficient to develop and research Arabic MT or design an Arabic MT system. Not even doing research in one of the related fields can achieve that. There must be whole-hearted cooperation between linguists, translators, computer engineers, and software programmers to achieve real progress. A computer

engineer may be a native speaker of Arabic, but that does not mean that he or she can do the job as well as a linguist. Only when there is interest, funding, and close cooperation between researchers in related fields, will there be a realistic chance for developing Arabic MT.

King Abdul Aziz City for Science and Technology (KACST). KACST is a governmental organization that conducts and supports research in all fields of science and technology. It realizes that translation is the only way to transfer scientific achievements and the latest technologies from all languages into Arabic (Alghamdi, 2009). It shows great interest in translation and in particular in MT.

MT is the field that brings together KACST's interest in translation and its interest in technology. Because it certainly is the case that MT, by being able to translate documents fast, can aid the transfer of knowledge and further new developments in science and technology. Therefore, KACST has a major role to play in MT, as explained by Alghamdi (2009):

- It launched a terminology database, called BASM, in 1982.
- It works on and funds research to develop algorithms, tools, and systems that help to improve MT and transliteration.
- It developed a free Arabic-English translation engine called AutoTranslator.
- It developed a system for automatic transliteration of Arabic names into English.
- It developed a system for automatic transliteration of foreign names into Arabic.
- It is working with IBM to develop an MT engine (Pangeanic, 2009).
- It funds research on MT, NLP, and Arabic computing conducted by universities, such as KSU and KFUPM, and is ready to fund any serious researcher in one of these fields.

KACST's members have conducted research on MT and other related fields. They seem to be ready to help and support researchers to improve the existing Arabic MT systems and even to build new and more developed ones.

Summary

In this chapter, the status of MT in Saudi Arabia has been analyzed. The analysis is based on data collected about the use of MT in Saudi governmental and nongovernmental organizations as well as translation agencies. The collected data also cover the teaching of and research into MT at language colleges and computer science and information technology colleges of Saudi universities. The analysis shows the following:

- About 60% of the translation agencies do not accept the idea of MT.
- Only 38.6% of Saudi organizations and 40% of translation agencies use MT, mostly Google Translate, which is used only as an online dictionary or for translating simple sentences.
- About 61.4% of Saudi organizations do not use MT, and 63% of them do not trust it at all.
- Only 45.5% of Saudi organizations, whether they currently use MT or do not use it, they plan to use MT in the future.
- Only 4.8% of the language colleges show interest in MT. No research is currently being conducted in any of these colleges.
- Only 25% of computer science and information technology colleges show interest in MT, and research is being conducted in only 17.9% of these colleges.
- Most studies were done by faculty of computing departments and no collaboration was taking place with linguists or translators.
- KACST is more engaged with MT than any other research center or university. It shows a keen interest in MT and conducts and supports research in this field and related areas.

Chapter 5: Conclusions

MT is not intended to replace translators or translation agencies. Translators will always be there and needed. However, simple texts or large volumes of texts of a certain technical nature can be translated by machines to enable translators to focus on texts that require their human creativity and intelligence. MT is helpful in translating texts for which only the gist is needed, or to decide whether it should be translated fully by a human translator. It is true that human translators, contrary to machines, can understand language nuances and culture, but they also forget, need time to remember, and get tired and make mistakes, even when they are highly experienced, try to be meticulous, and follow very strict rules. Machines, on the other hand, can work 24 hours, 7 days a week; they never forget, never get bored or tired, can be fed with millions and millions of pieces of information, and can produce the output in record time; however, they can break down.

This thesis has investigated the status of MT in Saudi Arabia. The purpose was to determine to what extent MT is used and how much interest in MT is shown by Saudi organizations, translation agencies, and universities and how much research is being conducted by Saudi universities and research centers on MT.

The study has shown that there seems to be relatively little interest in MT, which leads to very little research being conducted. The lack of interest is caused by the assumption that MT cannot produce accurate translations and, therefore, cannot be trusted. Most translators do not like the idea of a rough translation even if it can be improved by post-editing. Full post-editing can be more time-consuming and expensive than human translation. So, translators prefer to translate from scratch rather than post-edit MT outputs. Yet, if only rapid post-editing (or no post-editing at all) is needed, MT can save time and money. It is also shown that Arabic poses some difficulties for MT that need to be dealt with. Moreover, Sakhr is the only company in the Arab world that has an interest in MT, does research on MT, and has built Arabic MT systems. However, Sakhr is expensive for the Arab market. In addition, there appears to be no cooperation between language colleges and computer science colleges. Finally, there is no funding for research

in this field offered by the Saudi government. Only two universities, KFUPM and KSU, offer funding for research by their faculties. KACST is the only research center that supports and funds research in this field.

The findings of this study suggest that more attention to MT needs to be paid by Saudi governmental and nongovernmental organizations, translation agencies, and universities. More research needs to be conducted by universities and research centers. Interest will lead to more research, which, in turn, will help to improve the available Arabic MT systems and build more advanced ones. Therefore, there is a definite need for serious cooperation between faculties of both language colleges and computer science colleges. Linguists of Arabic and other languages, translators, computer engineers, and software programmers should all work together. Linguists must play their role in specifying the rules of the languages to be translated from and into, computer engineers and software programmers need to devote themselves to building the systems, and translators must evaluate the systems and the output quality.

Even though it is widely believed that machines cannot produce translations as accurately as humans do, MT systems can produce acceptable translations for certain text types, and if more research is conducted, then more improved Arabic MT systems will be built to produce more acceptable translations or to translate other types of text.

Bibliography

English

Abu Shquier, M. & Mohammed, T. (2008). *Agreement and word ordering with Arabic machine translation: A rule-based approach*. Paper presented at the Arab Conference on Information Technology (ACIT), December 15-17, 2008, Tunis.

Abu Shquier, M. (2010). *Evaluation of three English-Arabic MT systems against EA-RBMT*. Paper presented at Knowledge Management Information Conference (KMICe), May 25-27, 2010, University of Utara Malaysia, Kedah.

Abu Shquier, M., & Abu Shqeer, O. (2012a). Two experiments to evaluate the state of art of Arabic machine translation with EA-RBMT. Submitted for publication in *World Applied Sciences Journal (WASJ)*.

Abu Shquier, M., & Abu Shqeer, O. (2012b). Word ordering and corresponding verb-subject agreement in English-Arabic machine translation: An enhancement approach. Submitted for publication in *the International Arab Journal of Information Technology (IAJIT)*.

Abu Shquier, M., & Al Nabhan, M. (2010). *Rule-based approach to tackle agreement and word ordering in automated EAMT.* Paper presented at European, Mediterranean and Middle Eastern Conference on Information Systems (EMCIS), April 12-13, 2010, Abu Dhabi.

Abu Shquier, M., & Sembok, T. (2007a). *Handling agreement in machine translation from English to Arabic*. Proceedings of the 1st international conference on digital communications and computer applications (DCCA, pp. 370-381). Irbid: Jordan University of Science and Technology.

Abu Shquier, M., & Sembok, T. (2007b). *RBMT English-Arabic machine translation: A rule-based approach*. Paper presented at Conference of Learning International Networks Consortium (LINC), August 23, 2007, Amman.

Abu Shquier, M., & Sembok, T. (2008). *Direct rule-based Arabic MT system.* Paper presented at the International Scientific Conference (FMNS), June 2008, Blagoevgrad.

Abu Shquier, M., Al Nabhan, M., & Sembok, T. (2010). Adopting new rules in rule-based machine translation. In D. Al-Dabass, A. Orsoni, R. Cant, & A. Abraham (Eds.), *Proceedings of the IEEE 12th international conference on computer modeling and simulation* (pp. 62-67). Cambridge: CPS & IEEE Computer Society.

Abu Shquier, M., Sembok, T. & Mohammed, T. (2008). Word agreement and ordering in English-Arabic machine translation. In H. B. Zaman,T. Mohammed, T. Sembok, K. Rijsbergen, L. Zadeh, P. Bruza, . . . & M.Taib (Eds.), *Proceedings of the IEEE international symposium on information technology* (vol. 1, pp. 1-10). Kuala Lumpur: IEEE Computer Society.

Ajeeb. (n.d.). http: //www.ajeeb.sakhr.com

Alahmadi, H. (2010). *Building Arabic WordNet semantic-based dictionary.* (unpublished master's thesis). King Abdul Aziz University, Jeddah.

AlAjlan, A., AlKhalifa, H., & AlSalman, A. (2008). *Towards the development of an automatic readability measurement for Arabic language*. Paper presented at the 3rd International Conference on Digital Information Management (ICDIM), November 13-16, 2008, London.

Al-Barhamtoshy, H., Saleh, M., & Al-Kheribi, R. (2007). *Infrastructure for machine translation.* Paper presented at the 7th Conference on Language Engineering, December 5-6, 2007, Ain Shams University, Cairo.

Alghamdi, M. (2009). *KACST role in translation*. Paper presented at the 3rd Conference on Languages and Translation, December 28-30, 2009, Riyadh.

Alghamdi, M. (2010). Romanizing Arabic proper names: Saudi Arabia experience. In S. Izwaini (Ed.), *Romanization of Arabic names. Proceedings of the international symposium towards a transliteration standard of Arabic names: Challenges and solution* (pp. 142-146). Abu Dhabi: Ministry of Culture, Youth and Community Development.

Alghamdi, M. (2011). *KACST role in computing Arabic language*. Paper presented at Computational Linguistics Forum, November 28-29, 2011, Princess Nora University, Riyadh.

Alghamdi, M., & Muzaffar, Z. (2007). *KACST Arabic diacritizer*. Paper presented at the 1st International Symposium on Computers and Arabic Language, March 25-28, 2007, Riyadh.

Alghamdi, M., Alkharashi, I. & Alsughayr, I. (2007). *CERI contributions in Arabic language research*. Paper presented at Expert Meeting on Electronic Arabic Dictionary, June 11-13, 2007, Damascus.

Alghamdi, M., Al-Salman, A., Alhuqayl, K., & Alsubay, S. (2007). *A computerized system to romanize Arabic names*. Paper presented at the 1st International Symposium on Computers and Arabic Language, March 25-28, 2007, Riyadh.

Alghamdi, M., Muzaffar, Z., & Alhakami, H. (2010). Automatic restoration of Arabic diacritics: A simple, purely statistical approach. *Arabian Journal for Science and Engineering, 35,* 137-155.

Al-Haqil, A. (2010). *Machine translation versus human translation: A comparative study of selected texts.* (unpublished master's thesis). Imam Muhammad Bin Saud University, Riyadh.

Al-Hashim, A. (2009). *Arabic database for automatic printed Arabic text recognition research and benchmarking*. (unpublished master's thesis). King Fahd University of Petroleum and Minerals, Dhahran.

AlKhateeb, J., Ren, J., Jiang, J., & Al-Muhtaseb, H. (2011). Offline handwritten Arabic cursive text recognition using hidden Markov models and re-ranking. *Pattern Recognition Letters, 32*, 1081-1088.

Allen, J. (2003). Post-editing. In H. Somers (Ed.), *Computers and translation: A translator's guide* (pp. 297-317). Amsterdam: John Benjamins.

Al-Lihaidan, K. (2010). *Machine translation: An aid for humans or an indispensable tool for translators?* (unpublished master's thesis). Imam Muhammad Bin Saud University, Riyadh.

Almberg, E. (2002). Shall we dance, when the smart machines take over, virtually? In C. Sin-Wai (Ed.), *Translation and information technology* (pp. 185-193). Hong Kong: The Chinese University Press.

Almisbar. (n.d.). http://www.almisbar.com

Al-Muhtaseb, H. (2003). *An automated system to transliterate Arabic names into English: Towards a standard procedure*. Proceedings of the 1st workshop on Arabic name transliteration (pp. 143-174). Riyadh: KACST.

Al-Muhtaseb, H., & Aref, M. (1994). *A query language for Arabic expert system applications.* Paper presented at the 9th International Symposium on Computer and Information Sciences (ISCIS), November 7-9, 1994, Antalya.

Al-Muhtaseb, H., & Aref, M. (1996). Khabeer as a machine translation tool. In H. Al-Muhtaseb, & M. Aref (Eds.), *Proceedings of the 1st KFUPM workshop on information & computer science* (pp. 111-120). Dhahran: KFUPM.

Al-Muhtaseb, H., & Khayat, M. (1988). *Natural Arabic understanding system (NALUS).* Paper presented at the Regional Conference on Informatics and Arabization, March 9-11, 1988, Regional Institute for Informatics and Telecommunications, Tunis.

Al-Muhtaseb, H., & Mellish, C. (1997a). *From the generalized upper model towards an Arabic upper model*. Paper presented at the 4th IEEE International Conference on Electronics, Circuits, and Systems ICECS, December 15-18, 1997, Cairo.

Al-Muhtaseb, H., & Mellish, C. (1997b). *Towards an Arabic upper model: A proposal.* Paper presented at the 15th National Conference of Computers, November 17, 1997, Dhahran.

Al-Muhtaseb, H., & Mellish, C. (1998). *Some differences between Arabic and English: A step towards an Arabic upper model*. Paper presented at the 6th International Conference on Multilingual Computing, April 17-18, 1998, Cambridge, UK.

Al-Muhtaseb, H., Aref, M., & Al-Kulaib, A. (1994). *Khool: Khabeer object oriented language*. Paper presented at the 4th International Conference and Exhibition on Multilingual Computing, April 7-9, 1994, London.

Al-Muhtaseb, H., Aref, M., & Al-Mulhem, M. (1994). *English to Arabic machine translation: Goals, plans and steps.* Proceedings of the 1st symposium on computer applications (pp. 59-67). Manama: Bahrain University.

Al-Muhtaseb, H., Ashoor, M., & Khurshid, Z. (1994). *A step towards Arabic machine-readable cataloguing (ARABMARC).* Paper presented at the Conference on Exploiting Technology for Information Management in the Arabian Gulf Region, January 12-14, 1994, Bahrain.

Al-Muhtaseb, H., Elshafei, M., & Alghamdi, M. (2006). *Statistical methods for automatic diacritization of Arabic text.* Proceedings of the Saudi 18th national computer conference (vol. 18, pp. 301-306). Riyadh: Saudi Computer Society (SCS).

Al-Muhtaseb, H., Mahmoud, S., & Qahwaji, R. (2007). *Statistical analysis of Arabic text to support optical Arabic text recognition*. Paper presented at the International Symposium on Computer and Arabic Language & Exhibition (ISCAL), yNovember 10-12, 2007, Riyadh.

Al-Muhtaseb, H., Mahmoud, S., & Qahwaji, R. (2009a). A *novel minimal Arabic script for preparing databases and benchmarks for Arabic text recognition research*. Paper presented at the 8th WSEAS International Conference on Signal Processing *(SIP)*, May 30- June 1, 2009, Istanbul.

Al-Muhtaseb, H., Mahmoud, S., & Qahwaji, R. (2009b). A novel minimal script for Arabic text recognition databases and benchmarks. *The International Journal of Circuits, Systems and Signal Processing, 3*(3), 145-153.

Al-Qabbany, A., Al-Salman, A., & Almuhareb, A. (2009). *An automatic construction of Arabic similarity thesaurus*. Paper presented at the 3rd IEEE international Conference on Arabic Language Processing (CITALA), May, 4-5, 2009, Rabat.

Al-Rabiah, M. (2006). *An Arabic semantic parser and meaning analyzer*. (unpublished master's thesis). King Saud University, Riyadh.

Al-Rabiah, M., & Al-Salman, A. (2010). *An XML-based semantic parser for traditional Arabic*. Paper presented at the 4th IEEE International Universal Communication Symposium (IUCS), October 18-19, 2010, Beijing.

Al-Rumaih, D. (2010). *Machine translation vs. human translation in "The Secret" by Rhonda Byrne*. (unpublished master's thesis). Imam Muhammad bin Saud University, Riyadh.

Al-Salman, A. (1996). An Arabic programming environment. In K. M. George, J. H. Carroll, D. Oppenheim, & J. Hightower (Eds.), *Proceedings of the ACM symposium on applied computing* (pp. 480-486). Philadelphia, PA: ACM Press.

Al-Salman, A., Alohali, Y., & Alrabiah, M. (2006). An Arabic semantic parser and meaning analyzer. *Egyptian Computer Science Journal, 23*(3), 8-29.

Al-Sikhan, A. (1999). *Web page translation.* (unpublished master's thesis). King Saud University, Riyadh.

Al-Sikhan, A., Zantout, R., & Guessoum, A. (1999). *Automating the evaluation of machine translation systems lexicons: Arabic machine translation systems as case studies*. Paper presented at the 7th International Conference on Artificial Intelligence Applications, February 3-7, 1999, Cairo.

Al-Subaihin, A., Al-Khalifa, H., & Al-Salman, A. (2011). *A proposed sentiment analysis tool for modern Arabic using human-based computing*. Paper presented at IIWAS/MoMM Conference, December 5-7, 2011, Ho Chi Minh City, Vietnam.

Anas, T., & Alghamedi, M., (2011). *A corpus-based linguistics approach for estimating Arabic online content*. Paper presented at Conference on Human Language Technology for Development, May 2-5, 2011, Alexandria, Egypt.

AppTek. (n.d.a). [home page]. Retrieved November 28, 2011, from http://www.apptek.com

AppTek. (n.d.b). *TranSphere machine translation system*. Retrieved November 28, 2011, from http://www.apptek.com/index.php/transphere-machine-translation-system

AppTek. (n.d.c). *WebTrans machine translation system*. Retrieved November 28, 2011, from http://www.apptek.com/index.php/webtrans-web-based-machine-translation

Aramedia. (n.d.) *Sakhr Enterprise Translation (SET) V. 5.1*. Retrieved November 28, 2011, from http://www.aramedia.com/set.htm

Aref, M., & Al-Muhtaseb, H. (1993). *Khbeer: An Arabic expert system shell*. Paper presented at the 18th International Conference for Statistics, Computer Science & Social Applications, April 6-8, 1993, Cairo.

Aref, M., & Al-Muhtaseb, H. (1995). Khabeer: An Arabic object oriented production system and query language. *Processing Arabic Journal*, report 8, 77-105.

Aref, M., & Al-Muhtaseb, H. (1997). Khabeer: An object oriented Arabic expert system shell. *The Arabian Journal for Science and Engineering, 22*(2B), 275-293.

Aref, M., Al-Mulhem, M., & Al-Muhtaseb, H. (1995). English to Arabic machine translation: A critical review. In M. Fatani (Ed.), *Proceedings of the 4th Saudi engineering conference* (vol. 3, pp. 421-427). Jeddah: King Abdul Aziz University.

Arnold, D., Balkan, L., Meijer, S., Humphreys, R., & Sadler, L. (1994). *Machine translation: An Introductory guide*. London: NCC Blackwell.

ATA Software. (n.d.a). [home page]. Retrieved November 28, 2011, from http://www.atasoft.com

ATA Software. (n.d.b). *Al Wafi V. 4.0*. Retrieved November 28, 2011, from http://www.atasoft.com/goldenalwafi/main.htm

ATA Software. (n.d.c). *Almisbar online translation*. Retrieved November 28, 2011, from http://www.atasoft.com

ATA Software. (n.d.d). *Almutarjim Al Arabey V. 3.0*. Retrieved November 28, 2011, from http://www.atasoft.com/almutarjim/main.htm

ATA Software. (n.d.e). *Golden Al Wafi V. 2.0*. Retrieved November 28, 2011, from http://www.atasoft.com/goldenalwafi/main.htm

ATA Software. (n.d.f). *MutarjimNet V. 3.0.* Retrieved November 28, 2011, from http://www.atasoft.com/almutarjim/main.htm

Ba-Aziz, B. (2009). *Arabic language template grammars component based technology.* (unpublished master's thesis). King Abdul Aziz University, Jeddah.

Babylon. (n.d.a). [home page]. Retrieved April 28, 2012. from http://www.babylon.com

Babylon. (n.d.b). http://www.translation.babylon.com

Batawi, Y. (2011). *Evaluating the effect of N-Version programming technique on Arabic OCR accuracy: Experimental study.* (unpublished master's thesis). King Abdul Aziz University, Jeddah.

Bing . (n.d.a). *Bing translator*. Retrieved December 2, 2011, from http://www.microsofttranslator.com/help/?FORM=R5FD

Bing. (n.d.b). http://www.microsofttranslator.com

Bowker, L. (2002). *Computer-aided translation technology: A practical introduction*. Ottawa: University of Ottawa Press.

Bowker, L., & Barlow, M. (2008). A comparative evaluation of bilingual concordances and translation memories. In E. Yuste Rodrigue (Ed.), *Topics in language resources for translation and localisation* (pp. 1-22). Amsterdam: John Benjamins.

Chen, K., & Chen, H. (1996). A Hybrid approach to machine translation system design. *Computational Linguistics and Chinese Language Processing, 1*(1), 159-182.

Cimos. (n.d.a). *Al Nakel Al Arabi*. Retrieved November 28, 2011, from http://www.cimos.com/index.php?src=traduction

Cimos. (n.d.b). *Automatic translation.* Retrieved November 28, 2011, from http://www.cimos.com

Cimos. (n.d.c). *Multilingual Translation System (MLTS)*. Retrieved November 28, 2011, from http://www.cimos.com/index.php?src=traduction

Dilmanc. (n.d.). *Advantages and disadvantages of machine translation*. Retrieved November 20, 2011, from http://www.dilmanc.az/en/technology/mtadvantages

El-Affendi, M. (1987). *Implementation hints for the Arabization of programming languages*. Paper presented at the 1st KSU Arabization Symposium, April 6-9, 1987, Riyadh.

El-Affendi, M. (1991). An algebraic algorithm for Arabic morphological analysis. *The Arabian Journal for Science and Engineering, 16*(4B), 605-616.

El-Affendi, M. (1994). Sunbla: An intermediate step in a gradual promotion model for the development of natural Arabic programming systems. *The Arabian Journal for Science and Engineering (AJSE), 19*(3), 481-488.

El-Affendi, M. (1995). A *connectionist approach to Arabic morphological analysis.* Paper presented at the International Conference on Distributed Systems, March 1995, Kuwait University, Kuwait.

El-Affendi, M. (2002a). NeuroMorph: A connectionist MLP engine for Arabic morphological analysis. *The Egyptian Informatics Journal, 3*(1), 66-78.

El-Affendi, M. (2002b). On the morphological entropy of Arabic. *The Egyptian Computer Science Journal (ECS), 24*(2). Retrieved from http://dblp.uni-trier.de/db/journals/ecs/ecs24.html#El-Affendi02.

El-Affendi, M. (2008). A suggested framework for Arabic morphological analysis: A sliding window asymmetric matching algorithm and its implication. *The Egyptian Informatics Journal, 9*(1), 129-152.

El-Desouky, A., Abd-El-Gawad, A., & Saleh, M. (1996). A proposed algorithm for English-Arabic machine translation system. In H. Al-Muhtaseb & M. Aref (Eds.), *Proceedings of the 1st KFUPM workshop on information & computer science* (pp. 69-78). Dhahran: KFUPM.

European Association for Machine Translation (EAMT, n.d.). *What is machine translation?* Retrieved November 18, 2011, from http://www.eamt.org/mt.php

Farghaly, A., & Shaalan, K. (2009). Arabic natural language processing: Challenges and solutions. *ACM Transactions on Asian Language Information Processing, 8*(4), 1-22.

Global Security. (n.d.). *Machine translation techniques*. Retrieved November 20, 2011, from http://www.globalsecurity.org/intell/systems/mt-techniques.htm

Google Translate. (n.d.a). *Find out how our translations are created*. Retrieved December 2, 2011, from http://translate.google.com/about/intl/en_ALL/

Google Translate. (n.d.b). http://www.translate.google.com

Guessoum, A., & Zantout, R. (1998). *Towards a strategic effort, with a central theme of machine translation, to meet the challenges of the information revolution*. Paper presented at Symposium of Proliferation of Arabization and Development of Translation in the Kingdom of Saudi Arabia, September 23-24, 1998, King Saud University, Riyadh.

Guessoum, A., & Zantout, R. (2001a). A methodology for a semi-automatic evaluation of the language coverage of machine translation system lexicons. *The Journal of Machine Translation, 16*(2), 127-149.

Guessoum, A., & Zantout, R. (2001b). *Semi-automatic evaluation of the grammatical coverage of machine translation systems*. Proceedings of the MT Summit VIII (pp. 133-138). Santiago de Compostela: EAMT.

Guessoum, A., & Zantout, R. (2004). A methodology for evaluating Arabic machine translation systems. *The Journal of Machine Translation, 18*(4), 299-335.

Guessoum, A., & Zantout, R. (2006). Machine translation: A strategic dimension for the Arab world. *University Forum***, (**4), 32-37. University of Sharjah,

Guessoum, A., & Zantout, R. (2007). Arabic morphological generation and its impact on the quality of machine translation to Arabic. In A. Soudi, A.V. Bosch, & G. Neumann (Eds.), *Arabic computational morphology: Knowledge-based and empirical methods series: Text, speech and language technology* (vol. 38, pp. 287-302). Dordrecht, The Netherlands: Springer.

Habash, N. (2007). Arabic morphological representations for machine translation. In A. Soudi, A. Van den Bosch, & G. Neumann (Eds.), *Arabic computational morphology: Knowledge-based and empirical methods* (pp. 263-285). Dordrecht, The Netherlands: Springer.

Hamandi, L., Damaj, I., Zantout, R., & Guessoum, A. (2006). *Parallelizing Arabic morphological analysis: Towards faster Arabic natural language processing systems.* Proceedings of the conference on current issues in business and information technology (CIBITIC; pp. 455-459). Beirut: Haigazian University.

Hamandi, L., Zantout, R., & Guessoum, A. (2002a). *A Parser for the Arabic language*. Paper presented at the 16th Annual Symposium on Arabic Linguistics, March 1-2, 2002, Cambridge University, Cambridge.

Hamandi, L., Zantout, R., & Guessoum, A. (2002b). *Design and implementation of an Arabic morphological analysis system*. Paper presented at the International Conference on Research Trends in Science and Technology, March 4-6, 2002, Lebanese American University, Beirut.

Hannouna, Y. (2011). *Essentials of Arabic machine translation evaluation: The potentials and limitations*. Saarbrücken, Germany: VDM Verlag Dr. Müller.

Harrag, F., Al-Qawasmah, E., & Al-Salman, A. (2010a). A comparative study of statistical feature reduction methods for Arabic text categorization. In F. Zavoral, J. Yaghob, P. Pichappan, & E. El-Qawasmeh (Eds.), *Proceedings of the 2nd international conference on networked digital technologies* (NDT; pp. 676-682). Prague: Springer.

Harrag, F., Al-Qawasmah, E., & Al-Salman, A. (2010b). Comparing dimension reduction techniques for Arabic text classification using BPNN algorithm. In A. Ayesh, P. Pichappan, & A. Kumar (Eds.), *Proceedings of the 1st international conference on integrated intelligent computing* (ICIIC; pp. 6-11). Bangalore: IEEE Computer Society.

Harrag, F., Al-Qawasmah, E., & Al-Salman, A. (2011). *Stemming as feature reduction technique for Arabic text categorization*. Proceedings of the IEEE 10th international symposium programming and systems (ISPS; pp. 128-133). Algiers: IEEEXplore.

Harrag, F., Al-Salman, A., & Ben Mohammed, M. (2010). *A comparative study of neural networks architectures on Arabic text categorization using feature extraction.* Paper presented at the 1st International Conference on Machine and Web Intelligence (ICMWI), October 3-5, 2010, Algiers.

Harrag, F., Hamdi-Cherif A., & Al-Salman, A. (2009). *Applying topic segmentation algorithms on Arabic language*. Paper presented at the 5th International Conference for Computer Science Practice in Arabic, May 10-12, 2009, Rabat.

Harrag, F., Hamdi-Cherif A., Al-Salman, A., & Al-Qawasmeh, E. (2009). *Experiments in improvement of Arabic information retrieval.* Proceedings of the 3rd IEEE international conference on Arabic language processing (CITALA; pp. 71-81), Rabat: IEEE Computer Society.

Harrag, F., Hamid-Cherif A., & Al-Salman, A. (2010). Comparative study of topic segmentation algorithms based on lexical cohesion: Experimental results on Arabic language. *The Arabian Journal for Science and Engineering*, *35*(2C), 183-202.

Harrag, F., Hamid-Cherif A., Al-Salman, A., & Al-Qawasmah, E. (2011). Evaluating the effectiveness of VSM model and topic segmentation in retrieving Arabic documents. *The International Journal of Computer Systems Science and Engineering*, *26*(1), 55-68.

Holmes, J. S. (2000). The name and nature of translation studies. In L. Venuti (Ed.), *Translation studies reader* (pp. 172-185). New York: Routledge.

Homiedan, A. (1997). Machine translation. *King Saud University Periodical, Languages and Translation, 10*(1).

Homiedan, A. (1998). Machine translation. *Journal of King Saud University, 10*, 10-12.

Homiedan, A. (1999, July 27). *Analysis, transfer and generation in interlingua machine translation systems.* Language and translation lecture. Washington, DC: Georgetown University.

Homiedan, A. (1999, July 29*). Contribution of translatology to the development of 3rd wave machine translation systems*. Language and translation lecture. Washington, DC: George Mason University.

Hutchins, J. (2001). *Towards a new vision for MT.* An introductory speech presented at MT Summit Conference, September 18-22, 2001, Santiago de Compostela.

Hutchins, J. (2002). Machine translation today and tomorrow. In G. Willée, B. Schröder, & H. Schmitz (Eds.), *Computational linguistics: Achievements and perspectives* (pp. 159-162). Sankt Augustin: Gardez! Verlag.

Hutchins, J. (2003). Has *machine translation improved? Some historical comparisons.* Proceedings of the 9th machine translation summit (pp. 181-188). New Orleans, LA: AMTA.

Hutchins, J. (2005). Current commercial machine translation systems and computer-based translation tools: System types and their uses. *The International Journal of Translation, 17*(1-2), 5-38.

Hutchins, J. (2011). Recent applications of machine translation. In K. Malmkjaer, & K. Windle (Eds.), *The Oxford handbook of translation studies* (pp. 441-449). Oxford: Oxford University Press.

Hutchins, J., & Somers, H. (1992). *An introduction to machine translation*. London: Academic Press.

Izwaini, S. (2006). *Problems of Arabic machine translation: Evaluation of three systems*. Proceedings of the international conference on the challenge of Arabic for NLP/MT (pp.118-148). London: The British Computer Society (BSC).

Jambi, K., Saleh, M., & Al-Barhamtoshy, H. (2004). Language identification in document analysis (LIDA). In M. H. Rashid (Ed.), *Proceedings of the IASTED circuits, signals, and systems* (pp. 278-283). Clearwater Beach, FL: IASTED/ACTA Press.

Kenny, D. (2011). Electronic tools and resources for translators. In K. Malmkjaer & K. Windle (Eds.), *The Oxford handbook of translation studies* (pp. 464-472). Oxford: Oxford University Press.

Khayat, M., & Al-Muhtaseb, H. (1988). *Knowledge representation in natural Arabic understanding system. Proceedings of the 10th national conference on computers* (pp. 667-677). Jeddah: King Abdul-Aziz University.

Khuja, N. (2010). *Machine translation of Arabic compound words.* (unpublished master's thesis). King Abdul Aziz University, Jeddah.

Kit, C., Pan, H., & Webster, J.J. (2002). Example-based machine translation: A new paradigm. In S.W. Chan (Ed.), *Translation and information technology* (pp. 57-78). Hong Kong: Chinese University of Hong Kong.

Lawal, I. (2010). *Recognition of handwritten Arabic (Indian) digits using abductive network.* (unpublished master's thesis). King Fahd University of Petroleum and Minerals, Dhahran.

Lawson, V. (1988). A translator's map of machine translation. In M. Vasconcellos (Ed.), *Technology as a translation strategy, 2,* 105-115. Binghamton, NY: Center for Research in Translation, SUNY at Binghamton.

Lehrberger, J., & Bourbeau, L. (1988). *Machine translation: Linguistics characteristics of MT systems and general methodology of evaluation*. Amsterdam: John Benjamins.

Melby, A. (1987). On human-machine interaction in translation. In S. Nirenburg (Ed.), *Machine translation theoretical and methodological issues* (pp.145-154). Cambridge, UK: Cambridge University Press.

Munday, J. (2001). *Introducing translation studies: Theories and applications.* London: Routledge.

Nirenburg, S. (1989). New developments in knowledge-based machine translation. In J. E. Alatis (Ed.), *Georgetown University round table on languages and linguistics*, (pp. 344-357).

Obeid, H., & Zantout, R. (2007). *Line processing: An approach to ALPR character recognition*. Paper presented at ACS/IEEE International Conference, May 13-16, 2007, Amman.

Osman, Z., Hamandi, L., Zantout, R, & Sibai, F. (2009a). *Arabic optical character recognition.* Paper presented at the 10th Annual UAE University Research Conference, April 13-16, 2009. United Arab Emirates University, Al-Ain.

Osman, Z., Hamandi, L., Zantout, R., & Sibai, F. (2009b). *Automatic processing of Arabic text*. Proceedings of the international conference on innovations in information technology (IIT, pp. 140-144). Beirut: Beirut Arab University.

Othman, E., Shaalan, K., & Rafea, A. (2003). *A chart parser for analyzing modern standard Arabic sentences.* Proceedings of the MT Summit IX workshop on machine translation for Semitic languages: Issues and approaches (vol. 1, pp. 1-8). New Orleans, LA: AMTA.

Pangeanic. (2009, November 18). For the advancement of Arabic/English machine translation technology (and others): IBM and KACST [Web log post]. Retrieved from http://blog.pangeanic.com/tag/english/

Quah, C. K. (2006). *Translation and technology*. New York: Palgrave Macmillan.

Raído, V. E., & Austermühl, F. (2003). Translation, localization, and technology: Current developments. In L. Pérez González (Ed.), *Speaking in tongues: Languages across contexts and users* (pp. 225-248). València: Universitate de València.

SAIC. (2010). *SAIC expands human language technology offerings for federal and commercial customers*. Retrieved from http://investors.saic.com/phoenix.zhtml?c=193857&p=irol-newsArticle&ID=1495150&highlight

Sakhr Software. (n.d.a). *Arabic machine translation*. Retrieved November 28, 2011, from http://www.sakhr.com/mt.aspx

Sakhr Software. (n.d.b). [home page]. Retrieved November 28, 2011, from http://www.sakhr.com/Default.aspx

Salameh, M., Zantout, R., & Mansour, N. (2011). Improving the accuracy of English-Arabic statistical sentence alignment. *The International Arab Journal of Information Technology (IAJIT), 8*(1), 171-177.

Schäler, R., Way, A., & Carl, M. (2003). Example-based machine translation in a controlled environment. In M. Carl & A. Way (Eds.), *Recent advances in example-based machine translation* (pp.83-114). Dordrecht, The Netherlands: Kluwer Academic Publishers.

SDL TRADOS. (n.d.). *Translation memory*. Retrieved December 18, 2012, from http://www.trados.com /en/language-service-providers/translation-memory.asp

Sieny, M. (1985). *An Arabic dictionary for computer applications*. Paper presented at the 7th summer session (Informatics and Applied Arabic Linguistics), July 1985, of the Arab School of Science and Technology, Damascus.

Sieny, M. (1986). BASM: The Saudi terminology data bank. In M. Sieny & A. Wahab (Eds.), *Studies on machine translation* (pp. 198- 219). Riyadh: KACST.

Sieny, M. (1989). Machine translation in Saudi Arabia. *Journal of King Saud University, Language and Translation*, 143-145.

Somers, H. (2011). Machine translation: History, development and limitations. In K. Malmkjaer & K. Windle (Eds.), *The Oxford handbook of translation studies* (pp. 427-440). Oxford: Oxford University Press.

SYSTRAN. (n.d.a). [home page]. Retrieved November 26, 2011, from http://www.systransoft.com/systran

SYSTRAN. (n.d.b). *SYSTRAN Business Translator 7*. Retrieved November 26, 2011, from http://www.systransoft.com/translation-products/desktop/systran-7-business-translator

SYSTRAN. (n.d.c). *SYSTRAN Enterprise Server 7*. Retrieved November 26, 2011, from http://www.systransoft.com/translation-products/server/systran-enterprise-server

SYSTRAN. (n.d.d). *SYSTRAN Home Translator 7*. Retrieved November 26, 2011, from http://www.systransoft.com/translation-products/desktop/systran-7-home-translator

SYSTRAN. (n.d.e). *SYSTRAN Office Translator 7*. Retrieved November 26, 2011,, from http://www.systransoft .com/translation-products/desktop/systran-7-office-translator

SYSTRAN. (n.d.f). *SYSTRAN Premium Translator 7*. Retrieved November 26, 2011, from http://www.systransoft.com/translation-products/desktop/systran-7-premium-translator

SYSTRAN. (n.d.g). *SYSTRANBox.* Retrieved November 26, 2011, from http://www.systransoft.com /translation-products/online-services/systranbox

SYSTRAN. (n.d.h). *SYSTRANLinks*. Retrieved November 26, 2011, from http://www.systransoft.com /translation-products/online-services/systranlinks

SYSTRAN. (n.d.i). *SYSTRANNet*. Retrieved November 26, 2011, from http://www.systranet.com/translate

Tarjim. (n.d.). http://www.translate.sakhr.com/sakhr/MainView.aspx?lang=1)

The Language Translation. (n.d.). *Machine translation process*. Retrieved November 16, 2011, from http://www.thelanguagetranslation.com/machine-translation.html

Toury, G., (1995). *Descriptive translation studies and beyond.* Amsterdam: John Benjamins.

Trujillo, A. (1999). *Translation engines: Techniques for machine translation*. New York: Springer.

Tucker, A. B. (1987). Current strategies in machine translation research and development. In S. Nirenburg (Ed.), *Machine translation: Theoretical and methodological issues* (pp. 22-41). Cambridge, UK: Cambridge University Press.

Valconcellos, M.(1994). The issues of machine translation. In D.L. Hammond (Ed.), *Professional issues for translators and interpreters, 7,* 109-125. Amsterdam: John Benjamins.

Vauquois, B. (2003). Automatic translation: A survey of different approaches. In S. Nirenburg, H. Somers, & Y. Wilks (Eds.), *Readings in machine translation* (pp. 333-337). Cambridge, MA: MIT Press.

Zantout, R., & Guessoum, A. (2000). Arabic machine translation: A strategic choice for the Arab world. *Journal of King Saud University, Computer and Information Sciences, 12*, 117-144.

Zantout, R., & Guessoum, A. (2001). An automatic English-Arabic HTML page translation system. *Journal of Network and Computer Applications, 4*(24), 333-357.

Zughoul, M., & Abu-Alshaar, A. (2005). English/Arabic/English machine translation: An historical perspective. *Meta: Translators' Journal, 50*(3), 1022-1041.

Arabic

الحميدان، عبدالله. (1998) .الحاسوب والترجمة، ندوة تعميم التعريب وتطوير الترجمة ، 24 سبتمبر 1998، مركز الترجمة بجامعة الملك سعود، الرياض.

الحميدان، عبدالله. (1998). تطور نظم الترجمة الآلية، التواصل اللساني، مج 8، ع 1-2، (ص 5-31).

الحميدان، عبدالله. (1998). مفاهيم أساسية في الترجمة الآلية. نشرة بحثية، كلية اللغات والترجمة، جامعة الملك سعود.

الحميدان، عبدالله. (1999). التطبيقات الحاسوبية في الترجمة. محاضرة في المعهد العالي للغات الحية، 29 سبتمبر 1999، جامعة تونس الأولى للآداب والفنون والعلوم الإنسانية، تونس.

الحميدان، عبدالله. (2000). الطريقة التحويلية في الترجمة الآلية. المحاضرات الثقافية في إطار الجامعة والمجتمع،

21 أكتوبر 2000، جامعة الملك سعود، الرياض.

الحميدان، عبدالله. (2000). مفهوم الترجمة الآلية وتأثير البحث على المصطلح العلمي. محاضرة في النشاط الثقافي لكلية اللغات والترجمة، 21 نوفمبر 2000، جامعة الملك سعود، الرياض.

الحميدان، عبدالله. (2000). مقدمة في الترجمة الآلية. العبيكان للنشر.

الخياط ، محمد والمحتسب، حسني. (1987). نظام آلي لفهم اللغة العربية. ندوة جامعة الملك سعود الأولى عن تعريب الحاسبات، 6-9 أبريل 1987، الرياض.

الغامدي، منصور. (2006) مساهمة اللغويين العرب في مشاريع معهد بحوث الحاسب والإلكترونيات. أبحاث المؤتمر الوطني الثامن عشر للحاسب : تقنية المعلومات والتنمية المستدامة ، مج 18، ص (281-285)، الرياض.

الغامدي، منصور. (2006) نظام مقترح للنقل الكتابي بين أحرف جميع اللغات. ندوة: "النقل الكتابي بين اللغات: رومنة الأسماء العربية"، 13-15 نوفمبر2006، جامعة نايف العربية للعلوم الأمنية، الرياض.

الغامدي، منصور، محمد الكنهل، دحام العاني، فايز الحرقان. (2006) الترجمة عبر الشبكة العالمية: نظام حاسوبي مقترح للترجمة من وإلى اللغة العربية. مجلة جامعة الملك سعود: اللغات والترجمة، مج 18، ص (47 - 71).

(صيني) صالح، محمود إسماعيل. (2003). تخزين الأسماء العربية المكتوبة بالحرفين العربي والروماني في قواعد البيانات ومعالجتها حاسوبيا. أبحاث ندوة توحيد معايير النقل الكتابي لأسماء الأعلام العربية، (ص 367-378)، جامعة الأمير نايف للعلوم الأمنية، الرياض.

(صيني) صالح، محمود إسماعيل. (2007). البنية التحتية للمعالجة الحاسوبية للغة العربية. الندوة الدولية الأولى عن الحاسب واللغة العربية، 10 نوفمبر 2007، مركز الملك فهد الثقافي، الرياض.

صيني، محمود إسماعيل. (1984). الترجمة الآلية ونقل المعلومات. المجلة العربية، ع 78، ص (78 - 79).

صيني، محمود إسماعيل. (1989). المعاجم في الترجمة الآلية. أبحاث الملتقى الرابع للسانيات العربية والإعلامية (ص 183-196)، مركز الدراسات الاقتصادية والاجتماعية، تونس.

صيني، محمود إسماعيل. (1991). الترجمة الآلية: إمكاناتها وحدودها. المجلة العربية للثقافة، ع 21، (ص -132 144).

صيني، محمود إسماعيل. (1992). استخدامات الحاسوب في الترجمة. نشرة مركز الترجمة بجامعة الملك سعود، ص. 11.

صيني، محمود إسماعيل. (1994). الحاجز اللغوي: الحاسوب والترجمة. مجلة عربيوتر ، ع 49، (ص 22-34).

صيني، محمود إسماعيل. (1994). الترجمة الآلية واللغة العربية. مجلة التواصل اللساني، مج 6، ع 1- 2، (ص 83-87).

صيني، محمود إسماعيل. (1996). الحاسوب والترجمة. مجلة التواصل اللساني ، مج 3، (ص 1- 91).
صيني، محمود إسماعيل. (1996). الترجمة الآلية. مجلة الفيصل، ع 239، (ص 30-31).
صيني، محمود إسماعيل. (1996). التطورات الحديثة في الترجمة الآلية. أبحاث ندوة التعريب والحاسوب، (ص 144-169)، الجمعية العلمية السورية للمعلوماتية ، دمشق.
صيني، محمود إسماعيل صالح. (1999). الحاسوب في خدمة الترجمة والتعريب 1/2. المجلة الثقافية، س 6 ، ع 2، (ص 68-71).
صيني، محمود إسماعيل صالح. (1999). الحاسوب في خدمة الترجمة والتعريب 2/2 .المجلة الثقافية، س 6، ع 21، (ص 72-76).
صيني، محمود إسماعيل. (1999). بنوك المصطلحات الآلية والمعاجم الإلكترونية. السجل العلمي أبحاث الندوة الثانية لتعريب الحاسوب، (ص 311-334)، جامعة الملك سعود، الرياض.
صيني، محمود إسماعيل والعاني، عبدالرزاق. (1986). أبحاث ندوة دراسات في الترجمة الآلية ، (ص 35-219)، مدينة الملك عبد العزيز للعلوم والتقنية، الرياض.

Appendices

Appendix A: Questionnaire Designed for Saudi Organizations

AUS | الجامعة الأميركية في الشارقة American University of Sharjah

Questionnaire

Machine Translation in the Arab World: Saudi Arabia as a Case Study

A study by:
Faten AlMutawa
Master's Student

This questionnaire is designed to find out how widely machine translation is used in different organizations in Saudi Arabia. Your participation, by answering this questionnaire, will be very helpful and much appreciated.

Machine translation means using computer software to translate texts from one language into another. There are translation software packages available that translate from and into Arabic such as Sakhr, Al Mutarjim Al Araby, Al Wafi, and TranSphere. There are also online translation sites such as Google Translate, Tarjim, WebTrans, and Bing Translate.

Name of Organization: _____

1- Do you need or perform translation in your organization work?
☐Yes ☐No

A- If **Yes**, what do you need translation for?
☐ Official documents ☐ Paperwork ☐ Reports
☐ E-mails ☐ Websites
☐ Other, please state the Purpose: _____

B- If **Yes**, how often do you need translation?
☐ Always ☐ Occasionally ☐ Rarely

C- If **Yes**, do you have a translation department?
☐ Yes ☐No

If **Yes**, how many translation staff do you have?

If **No**, do you have a translator?
☐ Yes ☐No

If **No**, who does the translation, if any?
☐ A translation agency ☐ A freelance translator
☐ A bilingual employee whose main duties are not translation

2- Do you use machine translation?
☐ Yes ☐ No

A- If **Yes**, do you use:
☐ Translation software. Which one? _____
☐ Online translation. Which one? _____

B- If **Yes**, what do you use machine translation for?
☐ Official documents ☐ Paperwork ☐ Reports
☐ E-mails ☐ Websites
☐ Other, please state the purpose: _____

C- If **Yes**, do you post-edit the machine translation output?
☐Yes ☐No

If **No**, why not?☐Only a gist is needed ☐No final version is needed

D- If **No**, why do you not use machine translation?
☐ You do not trust it ☐ You cannot afford it ☐ Texts are in hard copy
☐ Other: _____

3- Are you planning to use machine translation in the future?
☐ Yes ☐ No

4- Any other remarks: _____

Appendix B: List of Saudi Organizations Included in the Survey

Ministries:

1. Ministry of Agriculture (MOA)
2. Ministry of Communication & Information Technology (MCIT)
3. Ministry of Culture & Information (MOCI)
4. Ministry of Economics &Planning (MEP)
5. Ministry of Finance (MOF)
6. Ministry of Foreign Affairs (MOFA)
7. Ministry of Hajj (HAJJ)
8. Ministry of Health (MOH)
9. Ministry of Higher Education (MOHE)
10. Ministry of Labor (MOL)
11. Ministry of Petroleum & Minerals Resources (MOPM)
12. Ministry of Transportation (MOT)

Governmental Organizations:

1. Capital Market Authority (TADAWUL)
2. Central Department of Statistics & Information (CDS)
3. Civil Defense
4. Communication & Information Technology Commission (CITC)
5. Consumer Protection Authority (CPA)
6. Control & Investigation Board (CIB)
7. E-Government Program (YESSER)
8. General Authority of Civil Aviation (GACA)
9. Human Rights Commission (HRC)
10. Islamic Development Bank (ISDB)
11. King Abdul Aziz Center for National Dialogue (KACND)
12. Presidency of Meteorology & Environment (PME)
13. Riyadh Chamber of Commerce & Industry (RCCI)
14. Royal Commission for Jubail & Yanbu (RCJY)

15. Saline Water Conversion Corporation (SWCC)
16. Saudi Arabian Monetary Agency (SAMA)
17. Saudi Commission for Tourism & Antiquities (SCTA)
18. Saudi Food & Drug Authority (SFDA)
19. Saudi Fund for Development (SFD)
20. Saudi Geological Survey (SGS)
21. Saudi Industrial Property Authority (MODON)
22. Saudi National Guard (SANG)
23. Saudi Ports Authority
24. Saudi Standards, Metrology & Quality Organization (SASO)
25. Shura Council
26. Supreme Council for Petroleum Affairs (SCPMA)
27. Supreme Economic Council (SEC)

Banks:

1. National Commercial Bank (NCB)
2. Saudi American Bank (SAMBA)
3. Saudi British Bank (SABB)
4. Saudi Investment Bank (SAIB)

Newspapers:

1. Alhayat
2. Alwatan
3. Arabnews

Hospitals:

1. AlHabib Hospital
2. International Medical Center (IMC)
3. King Fahd Medical City (KFMC)
4. King Faisal Specialist Hospital & Research Center (KFSHRC)

Companies and Nongovernmental Organizations:

1. Aggad Investment Group (AICO)
2. Alarrab Contracting Company (ARRAB)
3. Al-Babtain Group
4. Alfaisaliah Group
5. Almarai
6. Altayyar Group
7. Arabian Drilling Company (Oilfield)
8. Bin Ogan Group
9. DALLAH
10. E.A. Juffali & Brothers
11. Elajou Group
12. Freyssinet Saudi Arabia Company Limited (FSA)
13. Gecat
14. Hajj Ali Hussein Ali Reza Limited
15. Ibn Mahfouz Group
16. Jarir
17. King Abdul Aziz City for Science & Technology (KACST)
18. King Abdul Aziz Foundation for Research & Archives (DARAH)
19. Kingdom Holding Company
20. Makkah Construction & Development Company (MCDC)
21. Marine Works Limited (HUTA Group)
22. Mass Media for Advertising
23. Mediterranean & Gulf Cooperative Insurance & Reinsurance Company (Medgulf)
24. Modern Plastic Technology (MPT)
25. Naizak
26. National Company for Cooperative Insurance (Tawuniya)
27. Petrochem Arabia
28. Petromin Lubrication Oil Company

29. Sahara Petrochemical Company (Sahara PCC)
30. Saudi Advanced Industries Company (SAIC)
31. Saudi Airlines
32. Saudi Arabia Refineries Company (SARCO)
33. Saudi Arabian Amiantit Company (Amiantit)
34. Saudi Arabian Oil Company (ARAMCO)
35. Saudi Automotive Services Company (SASCO)
36. Saudi Basic Industries Corporation (SABIC)
37. Saudi Bin Laden Group (SBG)
38. Saudi Cable
39. Saudi Electricity Company (SE)
40. Saudi E-Tabadul Company (Tabadul)
41. Saudi Industrial Services Company (SISCO)
42. Saudi Oger
43. Saudi Pharmaceutical Industries & Medical Appliances Corporation (SPIMACO)
44. Saudi Telecom (STC)
45. SAVOLA
46. Thimar Investment Company
47. Trading & Contracting Company (Latifia)
48. Vinnellarabia
49. Yamama Cement
50. Zamil Industrial Investment Company (ZIIC)

Appendix C: List of Saudi Translation Agencies Included in the Survey

1. Abdulrahim Al-Harbi Translation Office
2. Al-Amri Translation Office
3. Al-Arabi Net for Translation
4. Al-Bayan Translation Office
5. Al-Gabas for Translation
6. Al-Khatlan Translation Office
7. Allewa Center
8. Al-Sajini Translation Office
9. Al-Shanawani Translation Office
10. Atlas Translation
11. Dar Al-Thukair for Translation
12. Day Translations
13. Dina Translation & General Services
14. Dr. Basem Alim Translation Office
15. Googan Group for Translation
16. HTO – Translations
17. Husam-Addin Sadagah Office for Translation
18. International Certified Translation
19. Kadasah International Center for Translation
20. King Faisal University Translation Center
21. King Saud University Translation Center
22. LadyLingua for Translation, Language & Design Services
23. Motargim.net
24. Muhammad Nader Law & Translation Office
25. Nusaiba International Translation Office
26. Sails Translation Services
27. Smart Translation Office
28. Universal Summit Translation Center

29. Wisam Al-Sinidi Translation Office
30. Zuhair Group for Certified Translation

Appendix D: Questionnaire Designed for Saudi Higher Education Institutions

AUS | الجامعة الأميركية في الشارقة American University of Sharjah

Questionnaire

Machine Translation in the Arab World: Saudi Arabia as a Case Study

A study by:
Faten AlMutawa
Master's Student

This questionnaire is designed to find out how much Saudi universities are interested in teaching and researching in machine translation. Your participation, by answering this questionnaire, will be very helpful and much appreciated.

Machine translation means using computer software to translate texts from one language into another. There are translation software packages available that translate from and into Arabic such as Sakhr, Al Mutarjim Al Araby, Al Wafi, and TranSphere. There are also online translation sites such as Google Translate, Tarjim, WebTrans, and Bing Translate.

Name of University or College: _____

1- Do you offer a translation degree at your university?
☐ Yes ☐ No

A- If **Yes**, do you offer a course in machine translation?
☐ Yes ☐ No

B- If **Yes**, how often do you offer this course?
☐ Every semester ☐ Every year ☐ Other: _____

C- If **No**, it is because:
☐There is no interest in machine translation ☐There is no funding available☐ There is no professor who can teach it ☐ Other: _____

2- Is there any professor specializing or interested in machine translation at your university?

☐Yes ☐No

3- Is there, in your university, any research done by faculty on machine translation?
☐Yes ☐No

A- If **Yes**, how many? Please state whether it is a conference presentation, paper, or a book: _____

B- If **Yes**, is this research adequately funded?
☐Yes ☐No

C- If **No**, it is because:
☐There is no interest in machine translation ☐There is no funding available
☐Other: _____

4- Are there, in your university, any graduate projects, theses or dissertations done by students on machine translation?
☐Yes ☐No

If **Yes**, how many? Please state the kind of work and level of study: _____

5- Are you planning to offer a course in machine translation in the future?
☐Yes ☐No

A- If **Yes**, why?
☐There is interest among students ☐There is a plan to expand
☐There is available funding ☐All the above

B- If **Yes**, it is because there is a market demand ☐

If so, has the market been surveyed by you or by another party?
☐Yes ☐No

C- If **No**, why?
☐There is no interest among students ☐There is no plan to expand
☐There is no available funding ☐All the above

D- If **No**, it is because there is no market demand ☐

If so, has the market been surveyed by you or by another party?
☐Yes ☐No

E- Other remarks: _____

Appendix E: List of Language Colleges at Saudi Universities Included in the Survey

1. Dar Al Uloom University (DAU)
2. EFFAT University
3. Imam Muhammad Bin Saud University (IMAMU)
4. Islamic University (IU)
5. Jazan University (JAZANU)
6. Juof University (JU)
7. King Abdul Aziz University (KAU)
8. King Faisal University (KFU)
9. King Khaled University (KKU)
10. King Saud University (KSU)
11. Najran University (NU)
12. Northern Borders University (NBU)
13. Prince Salman Bin Abdul Aziz University (SAU)
14. Princess Nora University (PNU)
15. Qassim University (QU)
16. Taibah University (TAIBAHU)
17. Taif University (TU)
18. Umm AlQura University (UQU)
19. University of Dammam (UD)
20. University of Hail (UOH)
21. University of Tabuk (UT)

Appendix F: List of Computer Science Colleges at Saudi Universities Included in the Survey

1. AlMaarefa College (MCST)
2. Dar Al Uloom University (DAU)
3. EFFAT University
4. Fahd Bin Sultan University (FBSU)
5. Global University
6. Imam Muhammad Bin Saud University (IMAMU)
7. Jazan University (JAZANU)
8. Jubail University College (JUC)
9. Juof University (JU)
10. King Abdul Aziz University (KAU)
11. King Fahd University for Petroleum & Minerals (KFUPM)
12. King Faisal University (KFU)
13. King Khaled University (KKU)
14. King Saud University (KSU)
15. Najran University (NU)
16. Northern Borders University (NBU)
17. Prince Muhammad Bin Fahd University (PMU)
18. Prince Salman Bin Abdul Aziz University (SAU)
19. Prince Sultan University (PSU)
20. Princess Nora University (PNU)
21. Qassim University (QU)
22. Taibah University (TAIBAHU)
23. Taif University (TU)
24. Umm AlQura University (UQU)
25. University of Dammam (UD)
26. University of Hail (UOH)
27. University of Tabuk (UT)

28. Yanbu University College (YUC)