

Problems of Arabic Machine Translation: evaluation of three systems

Sattar Izwaini
Abu Dhabi University
s.izwaini@adu.ac.ae

Abstract

The paper describes the translations of three online systems: Google, Sakhr, and Systran, using two sets of texts (Arabic and English) as input. It diagnoses the faults and attempts to detect the reasons, trying to shed light on the areas where the right translation solution is missed. Flaws and translation problems are categorized and analyzed, and recommendations are given. The two modes of translation (from and into Arabic) face a wide range of common linguistic problems as well as mode-specific problems. These problems are discussed and examples of output are given. The paper raises questions whose answers should help in the improvement of MT systems. The questions deal with establishing equivalents, lexical environment, and collocation. Cases that triggered these questions are illustrated and discussed.

Key words: MT Evaluation, Machine Translation, Arabic, NLP

1. INTRODUCTION

Since it was developed, Arabic machine translation has been subject to description and evaluation (Chalabi 2001, Farghaly and Senellart 2003, Al-Salman 2004, Al-Wasiti, 2005). The latter suggests that human-aided machine translation (HAMT) is the best method. He criticizes machine translation (MT) as being inefficient, and presents three samples of MT translations (without indicating which system produced them). One of those translations is Shakespeare's sonnet *Shall I Compare Thee*. MT has never been meant to deal with poetry, and it should not be. Creative writing cannot be claimed to be the domain where MT can be used, and cannot be taken as a criterion to evaluate its performance.

As Arabic and English are distant languages from two unrelated families, machine translation is bound to face many problems in producing meaningful coherent translations between these languages. When evaluating the output of MT, the transferred meaning is the most significant point of focus. Semantics is a very important aspect in translation as a theory and application and thus it requires our utmost attention.

The present paper provides a detailed account of the linguistic problems the three systems under study face. It attempts to diagnose the faults, find out the reasons, and suggests solutions, trying to shed light on the areas where the right translation solution is missed. The paper raises some questions whose answers should help in the improvement of MT systems: how to make the system take into account the lexical environment and collocation as one way to determine the right words to choose, and how to cater for word connotation, reference and gender. Cases that triggered these questions are illustrated and discussed.

2. DATA

Two sets of texts are used as input for three online MT systems that provide English-Arabic and Arabic-English translation: Google,¹ Sakhr's Tarjim,² and Systran. English texts are 76 in number and Arabic texts are 46. The texts are taken from an introductory course in translation given to Abu Dhabi University translation major students in the Fall semester 2005-2006. These texts are to be translated as an in-class activity and home assignments. The course material is designed to have texts that evolve in size and complexity along the course of study. The course starts with very short sentences such as *come in!* and *help!* then longer diverse sentences, some are of general nature are introduced. By the end of the course source texts develop into paragraphs of specialized nature of different areas. Literary texts are excluded from the data used for the study.

3. DISCUSSION AND EVALUATION

The present discussion is divided into two areas: Arabic into English and English into Arabic translations. The two modes of translation have some common problems as well as mode-specific problems. The discussion will focus on the linguistic problems the three systems face in dealing with the SL input and in producing the TL output. Diagnosing and analyzing problems can help in having lexically and grammatically well-formed translations, and thus optimizing the meaning communicated by the system output. I will look at readability; how fluent the translations are, and the level of informativeness; whether the original message is clear in the translation, and how close it is to an ideal translation by a human translator.

Arabic originals and translations will be glossed (between brackets) and when necessary with tags showing morphological and grammatical properties. Translations will be labelled as (GL), (SR), or (SN) to indicate which system produced them, whether Google, Sakhr, or Systran respectively.

4. ARABIC INTO ENGLISH TRANSLATION MODE

In this mode, many texts are rendered into correct, full and coherent English translations. The levels of readability, informativeness, and even grammaticality of those good translations are very high. Sakhr takes the lead with 16 translations (34.7%), Google comes second with 15 translations (32.6%), and Systran follows with two texts only (4.3%). One example is

من العوامل التي ادت الى الاقبال على المجمعات السكنية ذات الخدمات المتكاملة، نوعية المعيشة في المدن العصرية.

(from the-factors that led to the-demand for the-complexes the-residential of the-services the-full, quality the-living in the-cities the-modern)

Of the factors that led to the demand for residential complexes with integrated services, the quality of living in modern cities. (GL)

4.1 Problems of Lexis

The major problem in the Arabic-into-English mode is the non-vocalization of Arabic words, which leads to the wrong choice of TL words. Other problems include inadequate lexicon, multiple meanings, connotation, and collocation. One significant observation, however, is deletion which will be discussed first.

4.1.1 Deletion

One might argue that in the MT context deletion is a technical problem rather than a linguistic problem, in that it is related to the inability of the system to read and recognize words. Nevertheless, it causes a linguistic problem in the output. The present paper is, after all, about MT linguistic problems caused by technical pitfalls.

There are some cases of deletion by Google, which is peculiar since it is content words what is deleted. Some words are dropped with no apparent reason for such a procedure. MT systems tend to transfer all words of the input, let alone content words, as we will see later in this section. There are 10 cases of deletion by Google (in 21.7% of its translations). If these have no entry in the lexicon, they should be transliterated. The following lexical items are deleted: المجتمع، فقط، جاء، أشارك، عدد، جاءت، ستظهر، وصل، تعليمية أم القيوين. In the following example, the word عدد (issue/number) is dropped:

مجلة تراث تصدر عددا خاصا حول صيد الصقور.
(magazine Turath publishes number special on hunting the-hawks)
Heritage magazine published a special on hunting hawks. (GL)

4.1.2 Non-vocalization

Non-vocalization is a major cause for mistranslations. Having no diacritics renders many words homographs, and thus it is difficult to determine which meaning to choose. For example, الحر can mean *the-heat* or *the-free*, مهمة can be *task* or the feminine form of *important*, فقد can mean *to lose* or a conjunction plus a particle, يعد can mean *to promise* or *to prepare*, and يعقد can mean *to establish* or *to complicate*. All these words were wrongly translated. Only considering established translations of other lexical items in the TL text can help in determining which meaning to select. Here is another example:

توقع الإعلان عن اختيار المستشار الألماني الجديد غدا.
(expecting the-announcement about selecting the-chancellor the-German the-new tomorrow)
Sign announcing a new German Chancellor tomorrow. (GL)

As no diacritics available, the system wrongly processed the word توقع, producing an incoherent text. In the example below two words confused the system: العالم can mean *world* or *scientist*, and تفتح can mean *she opens* or *blooming*. In both cases, the system chose the wrong translation:

سيدني صاحبة أجمل ميناء في العالم تفتح ذراعيها للعرب وتدعوهم لزيارتها.
(Sydney owner most beautiful port in the-world open her-arms for-the-Arabs and call-them to-visit-her)
Reigned me friend of beautiful port in the scientist her blooming of arms for the Arab and calls them for her visit. (SN)

Probably because the lexicon is not updated to include the city name, the system is misled by the form of the input. It analyzed the city name as سيد (master/mister) and ني (the letter ن + me). When this pronoun is attached to a verb, not a noun, ن is added to cater for the phonological environment, namely to insert a consonant between two vowels. Such matching between this pronoun and a noun plus the consonant is ungrammatical. The system should have what makes it exclude such analysis.

Because words have no diacritics to determine their meaning, systems are confused between a verb and a noun of the same root, or between the verb in the active and in the passive, e.g. قتل and قُتل. However, even when diacritics are shown on the word above

(قنل), which makes a difference in meaning, the Google system did not manage to translate it into the right structure.

Non-vocalization is particularly important with verbs in the past form that have a subject-denoting suffix; only the final vowel can make a difference what subject the sentence should have: I, you (singular, masculine or feminine), or she:

إذا ذهبت إلى أثينا فلا تطلب فنجان قهوة تركي، فقد يقذف بك إلى خارج المقهى.

(If you-go to Athens then-not ask cup coffee Turkish, may thrown you outside the-cafe)

If I went to Athens not request a cup of coffee Turki, you jump out of the cafe. (GL)

If she went to Athens then does not request a Turkish coffee cup, may throw you to outside the cafe. (SR)

Whereas the two systems of Google and Sakhr could not determine the right person category, Systran made Athens the subject in the translation.

Moreover, the lack of diacritics of the common proper name element بن (the-son-of), as well as not using the person's first name and last name as a guide, has led to being translated into *coffee* (SN). A rule is needed to help the system to exclude such a translation.

Another aspect of processing Arabic words is that some letters are written in some Arab countries in different ways, e.g. final ي and *hamza*. These can affect the way the system recognizes the letter and the word, thus producing a wrong output. In Egypt, the final ي is written without dots, which makes it having the same form of the final *alif* ا. For example, علي مريض (Ali is ill) is translated into *On the patients* (GL), *On a patient* (SR), and *On sick* (SN). All systems translated the name into the preposition probably because they were confused by the form with dots.

4.1.3 Inadequate Lexicon

Non-updated lexicon is a common and simple problem that should be overcome easily. Surprisingly, some basic words are not included in the dictionary as the table below shows.

Google	Sakhr	Systran
مصافحته، بحريق، سنسافر تركي، بخطواته، يطفئ،	None	3. الحياة، أم القيوين، بك، أنحائها، يخلو، سنجرى -1 وزير التجارة الخارجية، منظمة التجارة العالمية -2 4. طبيب نفسي، رئيس وزراء

TABLE 1: Words and phrases not included in the lexicon

Having the definite article, the future prefix, a preposition, or a pronoun attached to the forms of the SL words seems to have made the system transliterate them. Manipulating the input by removing some of these elements has shown that the system recognizes the SL words and can translate them. Also, while the definite form of 'prime minister' is correctly translated, the indefinite form is not, which indicates that the latter form is not included in the dictionary. The basic form plus any added element needs to be catered for in the dictionary or at some stage of the translation process.

Further, many Arabic names have meanings and thus would confuse the system. These can be included in the lexicon to have them transliterated instead of being

translated. In the example below, the proper name is translated, and because of its position, it plays the modifier of the noun *states*:

حذر وزير التجارة الخارجية المصري رشيد محمد رشيد الدول الغنية...
 (warned minister the-trade the-foreign the-Egyptian Rasheed Mohammed Rasheed
 wealthy states...)
 External minister the commerce warned the Egyptian rational Mohammed the
 rational states rich... (SN)

The same name above is translated by Sakhr as *Rosetta is Mohamed Rosetta*. This translation is triggered by the Arabic name of the Rosetta stone after the Egyptian town (in Arabic *Rasheed*) where it was found. Such a shortcoming can be avoided by making a link with the proper names on one or both sides of the word and use them as a filter for such erroneous translation. Google managed to translate the name correctly by transliterating it, probably because its dictionary does not have the meaning *Rosetta* for this entry.

Place names can also have meaning and there is a need to have a rule to avoid producing literal translation:

اندلاع حريق كبير بمركز "الواحة" التجاري.
 (eruption fire big in-center "al-Waaha" the-commercial)
 The eruption of a big fire in the commercial oasis center. (SR)

Other examples include *Heritage magazine* for مجلة تراث (SR) and *mother [aalqywyn]* for أم القيوين (SN). Furthermore, some names are non-Arab names transliterated in Arabic and their original should be retrieved. For the same reason of non-updated dictionary, the name فان غوغ (Van Gogh) is dealt with by translating the first part into the meaning of an Arabic word that has the same form, rendering it as *transient*, while transliterating the second part as [GwG] (SN).⁵

Another important aspect is the translation of expressions made of more than one constituent element. These should be translated into one word (or two if it is originally three), depending on the SL text and the most appropriate equivalent or term in the TL. Nevertheless, all the words are kept in translation, most likely due to regenerating all elements of the SL text in the output.

SL expression	System translation	Idiomatic translation
نسبة 25%	25% ratio (SR) (SN)	25%
جولة سياحية	toured tourist (GL) tourist tour (SR) touristic tour (SN)	tour
شريط وثائقي	a tape of a documentary (GL) documentary tape (SR) documentary tapes (SN)	documentary
فنجان قهوة تركي	a cup of coffee Turki (GL) Turkish coffee cup (SR)	Turkish coffee
طبيب نفسي	psychological doctor (GL) (SR) doctor the breaths (SN)	psychiatrist

TABLE 2: Arabic expressions, their MT translations, and suggested idiomatic translations

These are idiomatic yet elaborate Arabic expressions that need to be rendered into idiomatic expressions. Terms in particular, e.g. طبيب نفسي (psychiatrist) should be treated as one unit, so that they can be translated into the right TL term. There are other idiomatic expressions of more than one element that need to be integrated in the system to have them correctly translated, e.g. لا بد (must), لا شك/من دون شك (undoubtedly), لا مناص (inevitably), لاسيما (especially). In one translation by Google, لا بد is treated as two elements, rendering it as *can not* thus producing a totally wrong meaning.

There are other cases where a one-to-one translation is expected, but the systems opted for another translation. For example, وجودهما (their-existence) is translated into *downpour* (SN), فنجان قهوة (coffee cup) into *teacup* (SN), (conferences) into *major* (GL), and 81 جريحا (81 injured) into *81 victim* (SN). In the last example, the noun is single in the SL according to the rule of nouns modified by number expressions. The translation also rendered it in the singular, a case that requires some attention. On the other hand, فنجان قهوة تركي (Turkish coffee cup) is translated into *Turkish teacup* (SN). If *teacup* is considered as a translation of فنجان (cup) then قهوة (coffee) is deleted.

4.1.4 Multiple Meaning, Connotation, and Collocation

Another major problem is words with multiple meanings. Many Arabic words can have two or more overlapping meanings in English, and systems need to decide which one to choose. The following words are translated by the three systems mostly into inappropriate meanings for the output:

بطولة	bravery, championship, heroism
إعلان	announcement, advertisement, declaration, sign
قمة	climax, peak, summit, top
يؤكد	confirm, emphasize, stress, verify
قوة	force, power, strength
مركز	centre, position, rank, status
موقع	position, rank, site, status
رئيس	major, president, primary
أسهم	arrows, shares, stocks

This problem requires consideration and a solution. Choosing any of these meanings can greatly undermine the coherence of the translation and the communication of the message. To solve this problem, a good approach is probably to take into account the lexical surroundings and collocations in the SL text, as well as the produced TL lexical items and their collocations. For example, افتتاح or يفتتح are translated into *inaugurates* or *inauguration* respectively. The TL words collocate with *championship*, which is the translation that should be chosen for بطولة, and not any other alternatives listed above.

In some cases, homographs can lead to mistranslations with no consideration of the topic of the text. For example, لها (to her/it) is translated into *fun* (SN) in a text about the UN, and يحل (to solve) is translated into *beautified* to collocate with *prices* (SN). Again, establishing the lexical surroundings in the TL can help in excluding these translations.

Connotation is another issue that requires attention. In some translations, TL words that are selected by the system have connotations that the SL words do not have. In the example below, the Arabic من يخلو has a negative connotation in its association with

المصداقية (credibility). Choosing *free of* in the output has made it positive and *credibility* something negative:

السلطة الفلسطينية: خطاب شارون يخلو من المصداقية.
(the-authority the-Palestinian: speech Sharon lack from the-credibility)
The Palestinian Authority: Sharon's speech is free of the credibility. (SR)

A similar change of connotation can be seen in using *afflicted* to translate ادت (led):

من العوامل التي ادت الى الاقبال على المجمعات السكنية ذات الخدمات المتكاملة، نوعية المعيشة في المدن العصرية.
(from the-factors that led to the-demand on the-complexes the-residential of the-services the-full, quality the-living in the-cities the-modern)
From the laborers which the residential attendance on the collectors afflicted to self of the services integral, nature of the living in the modern cities. (SN)

The translation is far away from even giving a gist of the original text. This is because some key words are translated incorrectly, and hence they seem out of context in the TL version, e.g. *laborers*, *attendance*, and *collectors*. The lexicon seems to have an entry for a classical less used meaning of the plural of عامل (laborer) which denotes a group of animals that carry out labor (Al-Waseet Dicitonary 1972: 628, Lisan Al-Arab Dictionary 1999: 401, Al-Qamoos Al-Muheet Dictionary 1994: 1339) or one of two plural forms of a female worker (Al-Munjid Dictionary 2002: 531) where the second form, عاملات, is the more widely used one. The system chose this meaning instead of *factors*, and also confused المجمع to mean *complex* with the one that possibly means the machine that collects, though the first is more common.

Priority of common usage needs to be looked at. When the system chooses a TL word, rarely used words should come last in priority. In using *signed* to translate the less common موقع (the-signed), the system passed over all other more common (and appropriate) words such as *rank* or *position*. These should come first in the system priority:

الصين تتنافس ألمانيا على مرتبة المنتج العالمي الثالث للسيارات. والهند تتجه لاحتلال موقع السوق الفرنسية بحلول 2015.
(China competes Germany on rank the-producer the-international the-third for-the-cars. and India is-heading to-occupy position the-market the-French by 2015)
China competition of German on rank of the producer worldwide third for the cars. India faces for signed occupation the market French in solutions 2015. (SN)

Probably one way to eliminate irrelevant TL equivalents is to make the system decide first the context by checking the key words and the semantic field(s) of the text.

Another important aspect in having meaningful well-formed text that has cohesion is collocation. Lexical items need to be translated with the constituent elements rendered according to their co-occurrence with each other. For example, مركز تجاري is translated literally into *commercial center* (SR) (SN) which gives quite different meaning, whereas the established collocation in the TL is *shopping center* or *shopping mall*. Even *mall* could be a better rendition. Another example of collocation is translating شريط وثائقي من ثلاث حلقات (documentary of three episodes) into *documentary tapes from three rings* (SN). Obviously, *ring* does not collocate with *documentary* or *tape*.

Some Arabic verbs are elusive in that they should not be translated only after considering their collocations. For example, the first meaning of راح is *went*, but when it

co-occurs with another verb, say *يدرس* (to study), then it means to *start studying*. Also, *يتناول* means to *eat*, or *take* (tablets), but when its subject collocate is *كاتب* (writer), *ناقد* (critic), *مقالة* (article), *كتاب* (book), or *برنامج* (programme), and/or its object collocate is *موضوع* (topic), *سيرة* (career), *شخصية* (personality), then it means to *investigate*, to *study*, to *highlight*, or to *cover*.

Another verb that needs attention is *يعيش* (live, experience) and its past form *عاش*. It can also be used to mean *long live*. In this case, it is incorporated in a short text. Despite its lexical environment and the text being long enough to mean to *live* or to *experience*, this verb is translated by Google into this marked special case of 'long live'.

4.1.5 Miscellaneous

One cultural aspect of Arabic is that it uses constructions that literally mean 'friend of', 'mother of', and 'father of' to indicate ownership, a characteristic, or an attribute, and hence need to be translated accordingly:

سيدني صاحبة أجمل ميناء في العالم تفتح ذراعيها للعرب وتدعوهم لزيارتها.
(Sydney owner most-beautiful port in the-world open her-arms for-the-Arabs and call-them to-visit-her)
Sydney's most beautiful harbour in the world opens its arm to the Arabs, and invites them to visit. (GL)

The Google system rendered this expression into a possessive relation, whereas the other two systems translated it literally as *the owner of* (SR) and *friend of* (SN). Systran also translated *أم القيوين* into *mother [aalqywyn]*.

Another remark here concerns transliteration. The Arabic alphabet does not have *p* and *v*. Any transliteration of Arabic words should not use these letters except probably for borrowed words. In the following translation, transliterating the verb with *v* is peculiar:

سنسافر الى الفلبين غدا.
(we-will-travel to the Philippines tomorrow)
Sensavr to the Philippines tomorrow. (GL)

4.2 Problems of Grammar and Syntax

Grammar and syntax is another problem in this mode; words are treated as individual blocks arranged one after another, resulting in awkward and almost meaningless translations. Looking at the last example in the previous section, we can note that the word order of the output reflects to a large extent the original. Putting words together according to the SL syntax bound to yield an output that needs too much post-editing.

Other problems in this mode include tense and aspect, reference and gender, prepositions, the definite article, and coordinators and conjunctions.

4.2.1 Word Order

Generally speaking, sentences in Arabic can be either nominal (SVO) or verbal (VSO). Arabic SVO order translates much more smoothly than the VSO order as it is more or less corresponds to the English word order. When the SL text has a VSO order a problem of syntax can arise, as the translation tends to copy the SL text, producing an awkward TL text:

قال الرئيس الباكستاني برويز مشرف من شأن مصافحته رئيس الوزراء الاسرائيلي ارييل شارون خلال وجودهما في نيويورك لحضور قمة الأمم المتحدة.

(underestimate Pakistani president Pervez Musharraf from matter his-handshaking Israeli prime minister Ariel Sharon during their-stay in New-York to-attend summit the-united the-nations)

Reduced Pakistani President Pervez Musharraf will Msafhath Israeli Prime Minister Ariel Sharon during their stay in New York to attend the United Nations summit. (GL)

Apart from the need to insert an auxiliary and the indefinite article, the following translation of a simple sentence requires an important step of post-editing, namely switching the positions of the noun and adjective:

هذه قصة قصيرة.

(this is a short story)

This story short. (GL)⁶

4.2.2 Gender and Reference

The reference system of Arabic uses pronouns of two genders only; there is no neutral gender. This can result in translating some masculine pronouns in the singular, and feminine pronouns in the singular and plural that refer to inanimate into *he/him/his* and *she/her*, resulting in confusion and incoherence of the translation. In such a case they should be translated into *it* or *its*. The MT system needs to establish a relation between the pronoun and the noun to which it refers, and identify whether it is animate or not, so that a TL gender can be determined:

أسعار النفط قرب مستوياتها القياسية.

(prices the-oil near their-levels the-standard)

Prices of the oil near even her standard. (GL)

In the following example, three references are made: the documentary, the man, and the state. The translation refers to the documentary as *he*, to the man as *he*, and to the state as *it*. Thus, the translation is confusing as it uses the same reference for two elements:

تتبع أهمية هذا الوثائقي من كونه لا يرصد فقط حياة الرجل الذي بنى دولة الإمارات وأسسها ووحّدها بل يحكي قصصاً واقعية ...

(spring-V importance this the-documentary from it-be-MASC not observe only life the-man who built state the-Emirates and established-it-FEM and united-it-FEM, but it-MASC-narrates real stories...)

An importance comes this documentary from being he does not observe only the life of the man that built the United Arab Emirates country and established it and he united it, but he narrates realistic stories... (SR)

Moreover, the system provides an inanimate reference for animate in the example below, probably because it overcorrects its output or fails to detect the reference in the input:

سؤال: لماذا قتل فان غوغ نفسه؟ «جواب: لأنه كان عاجزاً عن قتل طبيبه النفسي دكتور غاشيه.

(question: why killed Van Gogh himself? answer: because-he was incapable of killing his-psychiatrist doctor Gachet)

Question: Why kill the Gog himself? "Answer: because it was unable to kill psychological doctor Dr. Gachet. (GL)

4.2.3 Wrong Analysis of Input

Wrong analysis is probably caused by non-identification of different formats of Arabic words, i.e. when they are attached to prepositions and pronouns, as well as non-identification of diacritics, which have a major role in differentiating meanings. One system seems unable to recognize the verb form with plural inflection:

سنسافر الى الفلبين غدا.
(we-will-travel to the-Philippines tomorrow)
Philippines will travel to tomorrow. (SN)

Also, the singular form of the subject is rendered as plural. Even the particle قد is not recognized:

لقد نجحت!
(I have succeeded)
We have succeeded! (GL)
To succeeded! (SN)

In the example below, instead of analyzing وصلت as a verb with feminine inflection, the system analyzes it as a noun and pronoun, providing a misleading translation:

عاش المجتمع المصري حالة من الإثارة السياسية طوال الأشهر الماضية، وصلت ذروتها خلال الفترة السابقة لانتخابات الرئاسة.

(lived the-society the-Egyptian state of the-excitement the-political throughout the-months the-past, reached its-climax during the-period the-prior to-elections the-presidency)

The Egyptian society lived case from the excitement political throughout the months last, her connection of peak during the previous period for elections of the presidency. (SN)

A wrong analysis (and synthesis) can also lead to a coherent, grammatically well-formed TL text, but a one that is a total mistranslation and thus communicates wrong information:

هل يمكن لزيادة الرواتب في القطاعين العام والخاص ان تحل ازمة ارتفاع الأسعار؟

System's translation: Is it possible to increase salaries in the public and private sectors to solve the crisis of rising prices? (GL)

My translation: Is it possible that the increase of salaries in public and private sectors will solve the crisis of rising prices?

There are two readings of the SL sentence: one by the system that the increase is proposed, according which the system produces the translation, and the second reading has background knowledge that the increase has already been granted. That is why the semantics is different. How can a user of the system who has no knowledge of the SL know whether the translation is correct? Will there be a need for a reviewer and editor all the time?

One problem in the analysis of the input that the Systran system seems to face is processing numbers and the nouns they modify:

200 مليار درهم الاستثمارات العقارية المتوقعة خلال السنوات الخمس المقبلة.
(200 billion Dirham the-investments the-real-estate the-expected during the-years the-five the-next)

200 one billion Dirham the land investments expected during the years the five next. (SN)

The system does not process 200 and together, but rather separately. Consequently, is translated into *one billion* next to the figure, which can be confusing.

4.2.4 Tense and Aspect

As Arabic nominal sentences with a noun phrase plus an adjectival or prepositional phrase have no auxiliary (verb to be) when in the present tense, the translation has to have such an auxiliary inserted in the right position. The process of insertion needs to be obligatory in the system:

هذه قصة قصيرة.
(this is a short story)
This story short. (GL)⁷

Furthermore, as English and Arabic systems of tense and aspect do not correspond fully to each other, the translation of the Arabic past tense in cases of reporting should be translated into the present perfect because a past tense can sound as if it is a narration:

جاءت البنات.
(have-come the-girls)
The girls came. (SN)

In addition, the past tense is used in Arabic in conditional sentences to indicate a probable case of condition. When it is translated as a past tense, the condition would signify unlikelihood or improbability in the TL. Therefore, its translation needs to cater to the meaning of the original and the structure of the produced English text:

إذا ذهبت إلى أثينا فلا تطلب فنجان قهوة تركي، فقد يقذف بك إلى خارج المقهى.
(If you-go to Athens then-not ask cup coffee Turkish, may thrown you to outside the-cafe)
If I went to Athens not request a cup of coffee Turki, you jump out of the cafe.
(GL)

4.2.5 Prepositions

Translation of prepositions is an elusive task, especially for a machine. It is one of the thorny issues in MT (see also AlSharaf et al. 2004, regarding French/Arabic MT). The systems need to render prepositions according to the output and not according to the input. In not doing so, a text can be rendered with an inappropriate TL preposition that is either awkward or meaningless.

Many verbs and nouns have specific prepositions that associate with them and are part of the cohesion of the text. For example, اختتم is translated into *concluded by* (GL) or *in* (SN), while it should be *with*. Also, the verbs جاء and وصل are followed by a preposition when there is a need to specify a place or a person. Nevertheless, the Sakhr system seems to have *came to* as the only equivalent even though there is no need to have it in the TL:

وصل المدير.
(the manager has-come)
The manager came to. (SR)

Other examples of producing the wrong TL preposition include

الجمعية العامة للأمم المتحدة...
(General Assembly of the United Nations)
General Assembly for the United Nations... (SN)

عاش المجتمع المصري حالة من الإثارة السياسية...
(lived the-society the-Egyptian state of the-excitement the-political)
The Egyptian society lived a state from the political excitement... (SR)

81 جريحا بحريق في المنطقة الصناعية في إمارة الشارقة.
(81 injured in-fire in the-zone the-industrial in Emirate al-Sharjah)
81 injured by a fire in the industrial area in Emirate of Sharjah. (SR)

Wrong analysis of input can also produce mistranslation of prepositions. In the following example, the system seems to process the phrase بينهم (among-them) as if made of the preposition بـ (translated as *in*), the pronoun هم (translated as *their*), and بين as an element that has no entry in the lexicon and thus it is transliterated. The analysis is wrong in that the pronoun should be translated into *them*, and what the system considered as two elements is actually one which is the preposition بين :

قُتِل ما لا يقل عن 19 فلسطينيا، من بينهم مسؤول عسكري في حركة المقاومة الإسلامية (حماس).
{killed no less than 19 Palestinian, from among-them official military in
movement the-resistant the-Islamic (Hamas)}
Killed what 19 Palestinian does not decrease about, from in their yen military
official in movement resistant Islamic (Hamas). (SN)

How can the system be made to process such a phrase and translate it into a correct TL phrase? Can the lexicon have all forms where prepositions are attached to pronouns? I think yes, it can. These forms are finite and can be included in the lexicon.

4.2.6 The Definite Article

The Arabic definite article need not to be always transferred into *the*. Many defined words should be translated as indefinite. Generic names can be definite in Arabic but have to be indefinite in English. Also, when titles are associated with names, they do not take the definite article in the TL though they do in the SL, e.g. الشيخ (the-sheikh) to be translated into *Sheikh*, الملك (the-king) into *King*, الرئيس (the-President) into *President*. Moreover, institution names, place names, countries and cities can have the article in Arabic, but need to be stripped of it when transferred into English:

من العوامل التي ادت الى الاقبال على المجمعات السكنية ذات الخدمات المتكاملة، نوعية المعيشة في المدن العصرية.
(from the-factors that led to the-demand on the-complexes the-residential of the-services the-full, quality the-living in the-cities the-modern)
From the factors that led to the interest in the housing complexes with the integrated services, the living kind in the modern cities. (SR)

The definite article in *the housing complexes* and *the modern cities* need not be included in the translation. It also has to be dropped from *the political excitement* in the following example:

عاش المجتمع المصري حالة من الإثارة السياسية...

(lived the-society the-Egyptian state of the-excitement the-political)
The Egyptian society lived a state from the political excitement... (SR)

4.2.7 Coordinators and Conjunctions

Arabic is known for its extensive use of coordinators and conjunctions even within paragraph boundaries. Arabic texts abound with **و** (and) in the beginning of sentences and even of paragraphs. Coordinators and conjunctions need not always to be translated. When translated, in some cases, they affect negatively the English text cohesion. These need to have a smart filter to decide how to deal with them. A default rule is to delete them at the beginning of a sentence.

Furthermore, coordinators and conjunctions can cause problems when attached to other particles, e.g. **فقد** which is a conjunction attached to a particle to mean roughly *hence...has/have* or *hence...might*,⁸ is conceived by the Systran system as a verb and therefore it is translated *to lose*. One way to deal with this homograph is to identify the word that follows it. If it is a noun, it is a verb meaning *to lose*. If the word next to it is a verb, then it is a conjunction attached to a particle.

5. ENGLISH INTO ARABIC TRANSLATION MODE

In this mode, a number of texts are translated with the message clearly transferred in the TL with quite minimal or no post-editing required. Actually, the output is full, correct, and readable translation of the original, and provides a much higher level than just a gist. The Sakhr system is remarkable in this respect with 24 semantically correct translations even of some longer complex texts, making 31.5% of the total number. Second comes Google with 11 translations (14.4%), and last is Systran with eight translations (10.5%), both are restricted to short simple SL texts. The following example is even grammatically correct and accurately renders number and gender:

Turkey and Romania culled thousands of birds and imposed quarantine zones to try to stop the spread of avian flu.

تخلّصت تركيا و رومانيا من الآلاف من الطيور و فرضتا مناطق الحجر الصحيّ. لمحاولة إيقاف انتشار إنفلونزا الطيور. (SR)

(get-rid-of-FEM Turkey and Romania from the-thousands from the-birds and imposed-FEM-DUAL zones the-quarantine the-health to-try stopping spreading flu the-birds)

However, there are many problems in the output of this mode. These are divided into lexical and grammatical problems.

5.1 Problems of Lexis

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

5.1.1 Addition and Deletion

Two important drawbacks that have been observed in the translations produced by Google are addition and deletion. These phenomena are alarming since this is a machine work that deals with a written text, and hence there is no apparent reason as to why the system adds or deletes words. Translations by Google include three additions that have no SL counterpart:

I will never forget.

لن انساه ايدا. (GL)

(no I-forget-him ever)

In other two cases the system adds the Arabic word for *Dollar* next to numbers even though there are other currency signs next to them:

Microsoft Corp said on Tuesday it would pay RealNetworks Inc \$761 million (£436 million) to settle a suit accusing the world's largest software maker of using its dominance to promote its own software.

مايكروسوفت الشركة يوم الثلاثاء انها ستدفع رياننتوروكس المحدوده 761 مليون دولار (£ 436 مليون دولار لتسوية دعوي تتهم العالم اكبر صانع البرمجيات باستخدام هيمنتها علي تعزيز البرامج الخاصه. (GL)

(Microsoft the-company day the-Tuesday that-it will-pay RealNetworks the-limited 761 million Dollar (£436 million Dollar to-settle suit accusing the-world largest maker the-software by-using its-dominance on enhancing the-software the-private)

When manipulating the input, the translation seems to be affected by the word *million*; it brings with it the TL word for *Dollar*. In the above translation the system also left *said* with no TL counterpart, i.e. it deleted it. This is one of sixteen cases of deletion, which is peculiar. They should be either mistranslated or transliterated if they are not found in the dictionary. This means that about 21% of the translations by Google have some missing key element. In the example below *was found* is not translated:

An Iraqi journalist and photographer working for The New York Times in Basra was found dead early Monday after being abducted from his home by a group of armed men.

عراقي ومصور صحفي يعمل لصحيفه نيويورك تايمز في البصره ميتا اوائل الاثتين بعد اختطافه من منزله علي يد مجموعه من المسلحين. (GL)⁹

(Iraqi and photographer journalist work-V for-newspaper New-York Times in Basra dead first-PL the-Monday after abducting-him from his-home on hand group of the-armed)

5.1.2 Inadequate Lexicon

As is the case with the Arabic/English mode, the lexicon is also not updated. Words and phrases seem not to be included in the lexicon, and thus are transliterated. The lexicon lacks proper names: persons, cities, countries, as well as organizations and nationalities. Surprisingly, some simple phrases, like *too much*, are not dealt with successfully. The system just transliterates it, which is a standard procedure for words that are not found in the lexicon:

Skin discolouration and pigmentation is just one of the effects of too much exposure to the harsh Gulf climate.

جلد [ديسكلوريشن] وصبغ فقط واحدة من التأثيرات من [توو موش] انكشاف إلى القاسية خليج مناخ. (SN)¹⁰

(skin [deescolorayshin] and paint only one-FEM from the-effects from [too mosh] exposing to the-harsh-FEM gulf climate)

Other problems that can be identified in this example will be discussed under their relevant subsections. Below is a table of the words and phrases that are transliterated.

Google	Sakhr	Systran (58 words and phrases)	
None	discipline, discolouration, lied, Shaikh.	50 th , <i>the indefinite article a</i> , Abu Dhabi, African, Ajman, Al Qaeda, anarchy, birthday party, capitalization, construction company, Corp, Corporate communication, dirhams, construction material, Court of Appeal, discolouration, fast-developing, fast-evolving, Emirates, Flow, fountain of youth, got, growing numbers, health risks, Islamic, jail sentence, let's, lied,	Medical Centre, Mohamed, million people, mum-to-be, non-evasive, North Korea, one hundred, previous year, property developer, quiz, real estate, Saudis, scam, second time, series, setting up, Shaikh, South Asia, Sultan, superstar, Suspends, Telecommunications Corporation, to, too much, two armed (men), UAE, UK, Undersecretary, up to, World Health Organisation.

TABLE 3: Words and phrases not included in the lexicon

The table shows that many common, simple words, phrases and names are not included in the lexicon, let alone basic grammatical elements, which is a drawback in the system. About 76% of the output of Systran has a word or phrase that is transliterated. Inadequate lexicon is an easy problem to solve; such words and place names can just be compiled and integrated into the system.

The second problem here is terminology. The Lexicon has to include the TL terminology that can be used as equivalent in its output; otherwise, the translation can be misleading. Two systems, namely Sakhr and Systran, do not have the Arabic equivalent of *discolouration*, so they just transliterate it. Moreover, as Sakhr successfully translated both *bird flu* and *avian flu* into their Arabic term, Systran's dictionary does not have the TL equivalent of those two terms and thus it provides mistranslations with awkward constructions. The first, *عصفور إنفلونزا*, does not include the Arabic superordinate for *bird*, but a hyponym, i.e. a special kind of bird. The second translation, *إنفلونزا طيري*, has an awkward structure.

Google, on the other hand, translates *avian flu* into the awkward structure *الانفلونزا طيري* (the-flu avian) and *bird flu* into *انفلونزا الدجاج* (flu the-chicken). Google also translates *dentist* into the general term of *طبيب* (physician). These SL terms should be dealt with and processed as one unit, and the dictionary should include their TL counterparts so that its output can be correct.

In establishing equivalents there is a need to adopt the more widely used, less confusing terms. In the example below, *tagging* is translated into *تعليم*, a term based on colloquial usage to mean attaching labels or tags. It is, however, a homograph of the standardized Arabic term for *teaching* and *education*. The TL equivalent is inappropriate for this entry. With 'community' next to it, the reading is 'teaching' rather:

Yahoo seems likely to take a lead in the podcasting market with its new Podcast search beta: it looks easy to use, and it's got tagging and community ratings features.

ياهو يرجح الاضطلاع بدور راعد في بودكاستينغ السوق الجديد بودكاست البحث بيتا : يبدو سهله الاستعمال، واصبح التعليم في المجتمع وسمات التقديرات. (GL)

(Yahoo likely undertaking role pioneer in bodcasting the-market the-new bodcast the-search beta: it-look easy the-use, and became teaching in the-community and features the-estimation)

Finally, I would like to make a few remarks regarding the lexicon. Idiomatic phrases should be included as one unit such as *come in!* and *let's go*. When the system treats the elements of these phrases separately, it can provide one-to-one TL equivalents that when put together do not give the right TL phrase, e.g. *come in!* is translated literally into تعال في (SR) and not the idiomatic أدخل (literally enter). Punctuation could be a guide to the system in this respect.

The Systran system seems to have a problem with some (not all) words in upper case. When the input is a newspaper heading words are transliterated because they are capitalized, e.g. *Suspends* and *Flow*. When they are changed into lower case, the system translates them, as is the case with other words. Spelling also confuses the system. For example, *Shaikh* is transliterated, but *Sheikh* is translated.

Other capitalized words need to be given special attention so that they have a different status and thus need not to be translated, e.g. *Crown Plaza* is translated literally into ميدان التاج (SR) or literally with wrong structure التاج ساحة (SN), or partly into المملكة بلازا (GL) where *Crown* is translated as 'kingdom'. One rule can be included in the system to make it deal differently with capitalized words that are not found in the lexicon, so as not to look them up in the dictionary, but rather to transliterate them.

5.1.3 Multiple Meaning, Connotation, and Collocation

Polysemy and choosing the right TL meaning is a thorny issue in MT. Words with more than one meaning can confuse the system, which leads to the wrong choice of a TL word. According to the lexical surroundings of the SL word in the example below, *capital* is not عاصمة (capital city), but رأس مال (capital money).

Abu Dhabi Commercial Bank (ADCB) to double capital and offer 25% to foreigners.

بنك أبو ظبي التجاري (بنك أبو ظبي التجاري) لمضاعفة العاصمة و عرض 25 % إلى الأجانب. (SR)
(bank Abu Dhabi the-commercial (bank Abu Dhabi the-commercial) to-doubling the capital-city and offering 25% to foreigners)

The lexical environment could be a helpful tool for the system to opt for the right choice; the text elements can help in determining which word to choose. The collocation pattern of the chosen TL word does not allow its association with the TL word for *doubling*. Interestingly, we can see that the lexicon of the system is so updated that it even expands the abbreviated name, which is a merit, but not recommended when the short version is next to the full version of the name.

Translating into the right TL word can be elusive when dealing with words with different meanings, for example *lying*. It can mean making a statement that it is untrue or putting one's body on a flat surface. The system opted for the second meaning, which is the wrong translation:

A White House official was arrested Monday on charges of lying to investigators and obstructing a federal inquiry.

مسئول البيت الأبيض أُعْتَقِلَ يوم الإثنين بتهمة الاستلقاء إلى المحققين و إعاقه تحقيق فدرالي. (SR)
(official the-house the-white was-arrested day the-Monday by-charge the-lying to the-investigators and obstructing inquiry federal)

The system output is an almost correct translation of the original with the exception of the wrong translation of *lying*, which renders the TL text incoherent and the message defective. Making a link with the TL words for which the system already settled can help in choosing the right meaning of *lying*.

Different senses of the same word can also cause a problem. For example, *promote* means to publicize, enhance, or aid to progress, and is usually used in the passive to mean raise to a higher position. It is found to be translated according to the latter meaning into ترقية. Thus, the TL meaning deviates from the first meaning, which is the correct relevant one, to the second meaning which is the incoherent translation. The passive voice is a restriction that can be used to help the system determine which meaning to use:

Microsoft Corp said on Tuesday it would pay RealNetworks Inc \$761 million (£436 million) to settle a suit accusing the world's largest software maker of using its dominance to promote its own software.

شركة مايكروسوفت قالت يوم الثلاثاء أنها ستدفع لريالنيٹوروكس إينك 761 مليون \$ (436 مليون جنيه) لتسوية قضية تتهم أكبر صانع برامج للعالم باستخدام سيطرته لترقية برامجه. (SR)

(the-company Microsoft said day the-Tuesday that-it will-pay to-RealNetworks enk 761 million \$ (436 million pounds) to-settle suit accusing largest maker the-software for-the-world by-using its-control to-promoting its-software)

Another example is *get*. The system provides well-formed translation, but it renders *get* as 'understand' because the constituent elements of the original are not taken into account:

Ex-Tyco Executives Get 8 to 25 Years in Prison.

يفهم مسئولو إكس-تيكو 8 إلى 25 سنوات في السجن (SR)

(understand officials ex-tyeko 8 to 25 years in prison)

The TL word that is chosen to translate *get* does not take 'years' as an object. Can a rule be developed to exclude such translation of the verb when it associates with 'years' or any other TL word that makes the collocation unacceptable?

As shown above, collocation is significant in opting for the right translation. The collocation pattern of TL words chosen by the system can be utilized to function as a filter that eliminates non-matching words. The MT output usually sounds 'silly/funny' because one word is translated without due consideration to the lexical surroundings. The chosen TL word is out of context and thus the overall meaning is undermined. For example, while *give up* can be used with both *smoking* and *nuclear weapons*, its translation should change depending on the object. However, the system uses the Arabic verb used for smoking when the text is about nuclear weapons:

The lead American negotiator with North Korea signed a statement of principles that committed North Korea to give up "all nuclear weapons and existing nuclear programs."

مفاوض الصدارة الأمريكي مع كوريا الشمالية وقع تصريحًا للمبادئ أن كوريا الشمالية ملتزمة للإقلاع عن كل الأسلحة النووية و البرامج النووية الحالية. " (SR)

(negotiator the-leading the-American with North Korea signed statement for-principles that North Korea the-committed to-give-up from all the-weapons the-nuclear and the-programs the-nuclear the-existing.)

Similarly, in the example below, the system opted for a meaning that does not collocate with the object in the TL. In English, hunting can be for people and animals, but in Arabic it ought to be translated according to the object of the action. When it is animals it is يصطاد and when it is people it is يلاحق or يطارد (both roughly mean to chase):

Police are hunting a gang who allegedly bought millions of dirhams worth of construction material with cheques that bounced.

تصطاد الشرطه عصابة التي ظاهرياً اشترت ملايين من قيمة الدراهم لمادة البناء بالشيكات التي تتططت. (SR)
(hunt the-police gang who apparently bought millions from value the-dirhams for-material construction with-the-cheques that jumped)

In the same example above, the notion of 'silly/funny' translation is illustrated by the literal translation of the verb *bounced*. The established TL translations for *bouncing cheque* can be used to translate this collocation. Only the Google system could provide a right Arabic translation for it.

Another example of the collocation role in MT is the verb *blame*. As is the case with *hunt* above, this verb translates into two different Arabic verbs depending on its object. It is يلوم if it is a person or human-related party, or يعزو if it is non-human. The system, however, selects the human-related meaning, making the translation awkward:

Police blame sharp rise in crashes on reckless driving during Ramadan.

تلوم الشرطه الزيادة الحادة في التصادمات على القيادة أثناء رمضان المتهورة. (SR)
(blame the-police the-rise the-sharp in the-crashes on the-driving during Ramadan the-reckless)

In some cases, however, although the collocation is acceptable, the TL word reflects an attitude and perspective that the original text does not have:

Government official says rents will come down in next two years.

يقول المسئول الحكومي أنّ الإيجارات ستتهار في السنتان القادمتان. (SR)
(say the-official the-governmental that the-rents will-collapse in the-years-DUAL the-next-DUAL)

While *come down* is positive, its translation تنهار (collapse) has an exaggerated character and an extreme outcome. The problem here seems to emanate from having a one-to-one rigid lexicon and no collocation patterns built-in to play a role in the looking-up process.

Moreover, connotation is very important in transferring information. In translation, some words need to be translated into their corresponding charged or neutral counterparts to help in communicating the same attitude or point of view. One translation rendered *gang* into 'group' (SN), neutralizing its charge in the TL.

5.2 Problems of Grammar and Syntax

The main problems related to grammar and syntax in the English-Arabic mode are word order, gender and reference, tense and aspect, prepositions, the definite article, reference, and wrong analysis of grammatical categories.

5.2.1 Word Order

As the verbal sentence (VSO) is the default order in Arabic or more widely used, not taking care of it can produce unnatural nominal sentences. On the other hand, in newspapers headings, book and article titles, the nominal order is the norm and thus

SVO order is the right translation. Nevertheless, Sakhr tends to produce Arabic verbal sentences and thus a newspaper heading is not translated in the right order, resulting in awkward translation:

Turkey bus blast kills 2 and injures 10.
 (SR) يقتل انفجار حافلة بتركيا 2 و يجرح 10.
 (kill blast bus in-Turkey 2 and injure 10)

This is not always the case, however. Some Sakher's translations just ignore the rule of starting with a verb for no apparent reason. Systran, on the other hand, adopts the nominal order all the way through, and thus its output in general is an exaggerated towards the non-default word order of Arabic.

In addition, copying the ST syntax can result in incoherent translation, in some cases violating Arabic grammar. In the following example, the output does not even come close to the 'gisting' level. The only processing the system did is in switching the positions of one noun and its modifier but with a mismatching gender:

Police blame sharp rise in crashes on reckless driving during Ramadan.
 (SN) حافظت لوم إرتفاع حادثة في تحطامات على متهورّة يقود أثناء رمضان.
 (I-maintained blaming rise-MASC sharp-FEM in destruction-PL on reckless-FEM he-drive during Ramadan)¹¹

There is a probably a case here of mistranslation or deletion of *Police* and addition of the verb *حافظت* (maintain). Establishing a link of association between *driving*, *reckless* and *crashes* can help eliminating the mistranslation of *crashes*.

Finally, a basic rule of Arabic grammar states that when the subject is indefinite and the predicate is a prepositional phrase, the sentence starts with the prepositional phrase. Such a rule needs to be integrated in the system, so that ungrammatical output can be excluded:

There is a bird on the tree.
 (SR) هناك طائر على الشجرة.
 (there bird on the-tree)

Taking the dictionary meaning of *there* as a location in other contexts seems to have led to adopting this translation. *There* here denotes existence of an undefined subject; it is not a place adverb as in *he is standing there* to be translated as such. This is a recycled usage that translation has helped in its getting access into Arabic. Besides, there is a leniency on the part of many professionals working in translation and media regarding adhering to the rules of Arabic. MT and NLP specialists probably adopt the same attitude and this can make such an output for them absolutely acceptable.

5.2.2 Gender and Reference

Arabic has two genders: masculine and feminine. It has no neutral gender. Masculine, in single and plural form, is the default in Arabic and should be resorted as the first choice of the system. Feminine should be chosen when specific elements in the input clearly identify the gender such as *she*, *girls*, *it* with reference to a feminine, or ideally nouns whose Arabic counterparts are feminine such as *company*, *newspaper*, and *embassy* as well as cities and some countries.¹²

However, the Systran system seems to have the feminine as its default gender in its translations. The overwhelming majority of Systran's translations of nouns and

adjectives are rendered into feminine, though the original text has no indication of the feminine gender of the noun:

I am tired.
 أنا تعبـة. (SN)
 (I tired-FEM)

Even when there is an indication that the gender has to be masculine, the system produces a feminine form:

Ali is a baker.
 علي خبازة. (SN)
 (Ali baker-FEM)

No clear reason seems to be behind choosing the Arabic feminine adjective for *baker* and *tired*. The system seems to be configured to opt for the feminine form in the first instance, thus producing an ill-formed and misleading translation. Below is another example where there is no indication of the gender of the subject, yet it is translated into the non-default gender of the TL:

Expatriate workers can now renew their labour permits until they reach the age of 65.
 عاملات مغتربة يستطيع الآن جدّدت إجازاتهن عماليّة إلى أن يبلغ هم العمر من 65. (SN)
 (workers-FEM expatriate-FEM can-SING-MASC now renewed permits-their-MASC labour-FEM until that reach they-MASC the-age of 65)

Apart from the wrong message delivered by the translation, this configuration results in gender disagreement between the subject, the verb, and the reference. Disagreement in gender and number also occurs even though there is an indication of the gender. Here are translations by the three systems of the same SL sentence with more or less the same problem:

Women are being groomed to take their place in a more diverse professional elite.
 But many of these women say that is not what they want.

المراه يعده اخذ مكانه في اكثر تنوعا نخبه مهنيه. لكن الكثير من هؤلاء النسوة لا نقول ما نريد. (GL)
 (the-woman he-prepare-him he-took place in more diverse elite vocational. But the-many of these women not we-say what we-want)

هيأت نساء يكون أن يأخذ مكانهم في أكثر نخبة متوّعة محترفة. غير أن يقول كثير من هذا نساء أن ليس ماذا هم يريدون. (SN)
 (prepared-FEM women be that he-take place-their-MASC in more elite diverse professional. But that say many of this-MASC women that not what they-want-MASC they-MASC)

النساء يُهذّمَنَ لأخذ مكانهم في صفوفة محترفة مختلفة أكثر. لكن كثير من هذه النساء يقلن أن ذلك ليس ما يريدونه. (SR)
 (the-women well-groomed-FEM-PL to-take place-their-MASC in elite professional diverse more. But many of this-FEM women say-FEM-PL that that not what they-MASC-want)

While a high percentage of Systran translations suffers such inconsistency (43%), the Google system has nine cases (12%), and Sakhr has the least translations with only three texts (4%) that have such a discrepancy.

Further, gender disagreement also includes some person categories, as is the case with Google's translation above and the example below:

We can sit down together to discuss ways of making the plan work.
 نحن يستطيع جلست إلى أسفل معا أن يتناقش طرق من يجعل الخطة عمل. (SN)
 (we he-can I-sat to down together that he-discuss ways from he-make the-plan work-N)¹³

Reference is very important in communicating the right meaning of the SL text. MT systems in some instances fail to transfer the right reference network of the original. In the following example, although the verb agrees with the subject, the meaning is totally different. The translation misses the reference signals, and consequently the wrong party is liable for payment:

In an unprecedented move, the UAE Labour Minister took a tough stand against a construction company, asking it to pay its workers within 24 hours and slapped it with penalties as well.

في خطوه غير مسبوقة، واخذ وزير العمل الاماراتي صعبه تقف ضد شركات البناء ، وطلب من العاملين الدفع في غضون 24 ساعه والصفع لها عقوبات ايضا. (GL)
 (in step not preceded, and took minister the-labour the-emirati difficult-FEM stand-V-FEM against companies the-construction, and asked from the-workers pay within 24 hours and slapping to-her penalties also)

Now, the default gender in Arabic has to be adopted with a rule for deviation when there is an indication of the feminine gender. There is also a need for a rule to make a match between the subject, the verb, and pronouns to have a clear reference in the translation; who did what to whom. In this way the translation is not confusing or misleading, and it can communicate a relatively clear message.

5.2.3 Wrong Analysis of Input

Wrong renditions in some cases seem to result from the wrong analysis of the elements of the SL input. One example is when a verb with a third person inflection is identified as the plural form of the corresponding noun, and a noun is analyzed as a verb:

Palestinian chief demands end to anarchy after Israeli pull out.
 نهاية طلبات الرئيس الفلسطينية للفوضوية بعد اسرائيلي يتحرك. (SR)
 (end-N demands-N-FEM the-president the-Palestinian-FEM to-anarchy after Israeli he-move)

It seems that the system parsed *demands* as a noun, and *Palestinian* as an adjective modifying the head noun, leading to a total mistranslation. At the same time, the noun *pull out* is identified as a verb, leaving the modifier 'Israeli' with no noun to modify. The result is an incoherent translation that lacks readability and informativeness.

Moreover, a verb can be recognized as a noun leading to have one element out of context:

The earthquake's death toll rises above 20,000 as international aid and rescue teams fly to Pakistan to help the countless thousands battling to survive the chaotic aftermath.
 يترفع ضحايا الزلزال عن 20,000 كمعونة دولية و ذبابة فرق الإنقاذ إلى باكستان لمساعدة الآلاف الكثيرة بحاربون النجاة من النتيجة الفوضوية. (SR)

(elevate-himself-above-trivial-things victims the-earthquake about 20,000 as-aid international and fly-N-INSECT rescue teams to Pakistan to-help the-thousands the-many fighting survival from the-result the-chaotic)

Apart from other drawbacks, like the wrong translation of the verb *rises*, the translation provided the TL word for the insect 'fly' not the corresponding TL verb, making the output incoherent. As long as the system deals with a written text, the elements of the input should be utilized to help excluding irrelevant meanings. There is a need to develop a rule or a facility to be incorporated in the system to utilize, at some phase in the translation process, the lexical surroundings to determine the word meaning.

Here is another example of wrong translation of the subject probably triggered by wrong analysis of input, resulting in the misrepresentation of the message:

They talked a lot about their life together.

(GL) وتحدثنا كثيرا عن حياتهم معا.

(and we-talked much about their-life together)

Now I will move to the problem of building up nouns and modifiers together, which can be related to the analysis of input. MT systems seem not to identify the modifier and the head noun of the SL phrases they deal with, and consequently they fail to translate them into the right structure characteristic of Arabic. Another reason could be that systems do not have the rules for constructing compounds in Arabic. Either way, this results in an awkward structure that minimizes the readability and intelligibility of the translation:

Skin discolouration and pigmentation is just one of the effects of too much exposure to the harsh Gulf climate.

جلد [ديسكلوريشن] وصبغ فقط واحدة من التأثيرات من [توو موش] انكشاف إلى القاسية خليج مناخ. (SN)

(skin [deesclorayshin] and paint only one from the-effects from [too mosh] exposing to the-harsh gulf climate)

Here the gloss might seem right according to the English syntax. The translation, however, just puts words next to each other according to their precedence in the SL text. The sequence of Arabic noun clusters and adjectives is totally different. The right translation gloss should read (climate the-gulf the-harsh).

The same problem of identifying the head noun and its modifier(s) can be seen in the following two examples:

1- Traditional Olympic Advertising will flop at the 2008 games in Beijing, an expert on Chinese advertising has warned.

a- (SR) تقليديّ أولمبيّ سيسقط الإعلان في الـ2008 لعبة في بكين، خبير في الإعلان الصينيّ قد حذر.

(traditional olympic will-fall the-advertising in the-2008 game in Beijing, expert in the-advertising the-Chinese has warned)

b- (GL) الاولمبي الاعلان التقليديه...

(the-olympic the-advertising the-traditional-FEM...)

2- Ajman police quiz traffic department officers over driving license scam.

a- (SR) تستجوب شرطة عجمان مسؤولي قسم المرور على قيادة خدعة الرخصة.
(question-V police Ajman officials department the-traffic on driving deception the-license)

b- (GL) ..رخصه القيادة الغش.
(...driving the-license cheating)

The system just arranges words in the TL according to their positions in the SL text. While they managed in processing similar cases successfully, the systems seem to be confused when having a noun compound modified by another noun or by an adjective. In Sakher's translation (2 a), instead of parsing *driving license* as a modifier of *scam*, the system parsed *driving* as a modifier of *license scam* and constructed the translation accordingly.

While the above translations by Google and Sakhr need a few steps of post-editing, Systran's translation of the same SL sentences need far more steps due to the large number of errors that include untranslated words and awkward word order.

Not recognizing the constituents of the input can also produce a mistranslation with a deviant message, leading to grave consequences:

An Indonesian Islamic militant vowed revenge after he was sentenced to death yesterday for plotting a deadly bombing at the Australian embassy.

اندونيسيا الاسلاميه المقاتله تعهد بالثأر بعدما حكم عليه بالاعدام بتهمة التامر امس الداميه في تفجير السفارة الاستراليه. (GL)

(Indonesia the-Islamic the-militant vowed with-revenge after sentenced-him by-execution with-conspiracy yesterday the-bloody in bombing the-embassy the-Australian)

The system seems unable to identify the adjective *Indonesian* and/or translated it as a noun. The danger in the translation is that it deviates the meaning of the SL sentence in that an individual in the input becomes a whole nation in the output. Would a disclaimer and a limitation of liability repair the damage and mitigate the consequences? (see also Gaspari 2004: 66, regarding web sites that have links to free on-line MT service to have their contents translated).

Similarly, the information transferred in the translation below is also changed:

An Iraqi journalist and photographer working for The New York Times in Basra was found dead early Monday after being abducted from his home by a group of armed men.

صحفيّ عراقيّ و مصوّر اللذان يعملان مع النيويورك تايمز في البصرة وُجداً ميّتين في وقت مبكر من الإثنين بعد الخطف من بيته بمجموعة رجال مسلّحين. (SR)

(journalist Iraqi and photographer who-DUAL work-V-DUAL with the-New-York Times in Basra found-DUAL dead-DUAL in time early of the-Monday after the-abduction from his-home by-group men armed)

The inaccurate analysis of constituent elements of the subject seems to have led to having two persons instead of one. Also, the sentence has a reference to a singular person which should contradict the analysis that there are two people involved. The reference here, بيته (his-home), is an indicator that can help the system to provide the correct translation. Another kind of deviation can be seen in the following example, but this time the whole meaning of the SL sentence is changed:

The man was killed by two armed men.
 رجل قتل رجلين مسلحين. (GL)
 (man killed men-DUAL armed-DUAL)

It seems that the passive structure is not parsed by the system as such, leading to this totally different message.

5.2.4 Tense and Mood

English and Arabic systems of tense and aspect do not have one-to-one correspondence. The Sakhr system is well-configured to cater for this issue. It translates the present perfect into the closest structure constructed of the past tense and the particle قد which, when associates with the past, signifies completion of action.

Another issue is that English sentences in some cases have two (or more) verbs with the same tense as a matter of time relevance. Arabic, on the other hand, reports one verb (or more) in a different tense (usually present) because it is rather considered as a fact, e.g. *He told me he was tired* should be translated by eliminating the past tense of the second verb: أخبرني أنه متعب (he told me he is tired). The systems seem not to process their translations accordingly. In the example below, the past tense of the verb خطط (to plan) should be translated into the present tense تخطط:

OPEC delegates said Monday that the group planned to allow its members to provide up to two million barrels a day of additional crude oil if the market needs it.

قال مندوبو منظمة أوبك يوم الإثنين أن الجماعة خطّطت للسّماح لأعضائها أن يزودوا حتّى مليونان برميلاً يوم نפט خام إضافي إذا احتاج السّوق له. (SR)

(said delegates organization OPEC day the-Monday that the-group planned to-allow to-its-members that provide up-to million-DUAL barrels day oil crude additional if need the-market to-it)

A couple of remarks are to be made here. First, the translation of *group* into جماعة has made it having some religious profile and should be مجموعة or even better منظمة. Second, *a day* should be processed as an adverb and hence to be translated as يومياً (daily). There is a need to develop rules to cater to these cases.

Another grammatical problem is mood. It seems that imperative is not recognized. Thus any instructions to the user stated in the SL text are expected not to be communicated clearly. One simple sentence is translated correctly by Sakhr except for the absence of the object, with wrong mood (and wrong meaning) by Systran, and mistranslated by Google:

Pass the salt, please.
 ناول الملح، من فضلك. (SR)
 (give the-salt, please)

Pass the salt, please.
 مررت الملح، رجاء. (SN)
 (I-passed the-salt, please)¹⁴

Pass the salt, please.
 الترخيص الملح، الرجاء. (GL)
 (licensing the-salt, please)

The system can also opt for a TL verb form that deviates from the original meaning. *Expanded* is translated into the transitive form of its Arabic counterpart, whereas it should be the reflexive form توسّع :

Fearing they could disappear in a society that has expanded beyond anything most of them imagined, UAE nationals are voicing their concerns through cinema.
 خشية أنهم يمكن أن يختفوا في مجتمع قد وسّع بعد أي شئى معظمهم تخيلت، مواطنو الإمارات العربية المتحدة هم الضنّبط مخاوفهم خلال السّينما. (SR)
 (fearing that-they can that disappear in society has expanded-TRANS after any thing most-of-them imagined, citizens United Arab Emirates they the-exact their-concerns through the-cinema)

However, the use of the reflexive form of the verb instead of the transitive form causes a confusion of meaning:

Parents withdraw children from school over textbook controversy.
 والدات ينسحبون أطفال من مدرسة على كتاب مدرسيّ جدال. (SN)
 (mothers withdraw-themselves-MASC children from school on book scholastic controversy)

In both cases, the existence or non-existence of the object makes the difference in the meaning of the verb.

5.2.5 Prepositions

As is the case with the Arabic/English mode, prepositions pose a problem in this mode. English into Arabic MT provides the wrong preposition most of the time. A preposition has different meanings according to the noun or verb next to it. The best way seems to establish a link between these verbs and nouns and their Arabic counterparts so that prepositions are not dealt with separately.

Some verbs and nouns have specific prepositions with which they collocate, and there is a need to establish a link between them so that they appear together in the translation. While hitting in Arabic is 'on' the face, the translation copies the SL preposition *in*:

A man hit his receptionist in the face with a phone.
 ضرب رجل موظف استقباله في الوجه بتليفون. (SR)
 (hit man staff his-reception in the-face with-telephone)

As the system processes prepositions and translates them anyway by providing a TL corresponding preposition with no regard to neighbouring words, the translation would have extra elements that affect the readability and grammaticality of the output:

...growing numbers of women...
 ...الأعداد المتزايدة للنساء... (SR)
 (growing numbers for women)

...too much exposure to the harsh Gulf climate.
 ...تعرّض كثير جداً إلى مناخ الخليج الصّعب. (SR)

In this case, the right preposition to use with this noun is *لـ*. Gloss is not provided here, because no English preposition would show the subtle difference. Even educated native speakers of Arabic nowadays are confused between *لـ* and *إلى*, and use them interchangeably.

5.2.6 The Definite Article

An important element in the well-formedness of a translation into Arabic is the use of the definite article *al*. Generic names, many place names, and titles have to have the article in Arabic, e.g. *police الشرطة* (the-police). In some cases, the system wrongly inserts this article though the original does not as the case is with the subject in the following example:

Government official says rents will come down in next two years.
 يقول المسئول الحكومي أنّ الإيجارات ستتهار في السنتان القادمتان. (SR)
 (say the-official the-governmental the-rents will-collapse in the-years-DUAL
 the-next-DUAL)

Here is another example where the definite article should be attached to the modifier only:

The UAE's marketing culture is...
 الامارات التسويق الثقافة... (GL)
 (the-emirates the-marketing the-culture)

The translation has an awkward construction duplicating the sequence of the elements of SL structure with unnecessary addition of the definite article to the head noun. In Arabic, when one noun modifies another noun, the head noun when definite must not have the definite article while the modifier must.

An important and simple grammatical rule is that verbs cannot be definite or indefinite. However, one translation does violate this rule only because the present participle is translated into a verb. If it is a modifier, it can have the definite article attached when the noun is definite as well. The translation, however, does not observe the rule:

Twilight in the desert: the coming Saudi oil shock and the world economy - By
 Matthew R. Simmons. \$24.95.
 غسق في الصحراء: ال يأتي سعودية زيت صدمة والاقتصاد عالمي - بماتيو [ر.] [سمونس]. \$24.95. (SN)
 (twilight in the desert: the he-come Saudi oil shock and the-economy international
 - with-Matthew [R.] [Smons]. 24.95\$)

A filter can be developed to help excluding such matching between the article and a verb. Another rule in Arabic grammar is that the definite article should not be attached to غير (roughly *non*) when it is part of an expression like غير مدفوع for *unpaid*:

More than one hundred construction workers blocked Shaikh Zayed Road on
 Monday morning demanding unpaid wages.
 أكثر من مئة عامل بناء سدّ طريق الشيخ زايد صباح الإثنين الذي يطلب الأجرة الغير مدفوعة. (SR)
 (more than one-hundred worker construction blocked road the-Sheikh Zayed
 morning the-Monday who-SING-MASC he-require the-wage the-non paid)

The system rightly inserted the *al* with the title *Shaikh* as titles that are indefinite in English have to be definite in Arabic. However, it added the article to a particle that cannot be definite, but rather the noun that follows it can. A grammatically correct translation should read غير المدفوعة (non the-paid). A rule can be formulated to have the system attach the definite article to the noun rather than to the particle.

5.2.7 Wh-words

These need to be correctly processed. They can be translated either as question words or relative pronouns. The syntax and punctuation can be set as a guide to the system to determine how to render wh-words. In the following example the wh-word is mistranslated as a relative pronoun:

Who are you?
الذي يكون أنت؟ (SN)
(the-one be you?)

The sentence structure and the question mark can guide the system that this is an interrogative sentence and a question word should be used in the translation.

In the example below, a relative pronoun is wrongly inserted. Only when there is a definite noun, can a relative pronoun be introduced:

An Iraqi journalist and photographer working for The New York Times in Basra was found dead early Monday after being abducted from his home by a group of armed men.

صحفيّ عراقيّ و مصوّر اللّذان يعملان مع النيويورك تايمز في البصرة وُجداً ميّتين في وقت مبكر من الإثنين بعد الخطف من بيته بمجموعة رجال مسلّحين. (SR)

(journalist Iraqi and photographer who work with the-New-York Times in Basra found dead in time early of Monday after the-abduction from his-home by-group men armed)

At the end of this sub-section, I would like to make a final remark about the grammaticality of translations. Although not abiding by case endings in the Arabic output will not affect the meaning much, the output is not a spoken language and thus the translation needs to observe the Arabic system. After all, it is the case system that makes it clear who is doing what to whom and not the position of the words in the sentence.

The TL parsing should also be maintained in the output. A verb preceded by the particle حتى takes the subjective mood, and if it is inflected for the plural, its final ن has to be dropped. The translation below does not follow the above-mentioned rule although it is almost perfect:

Expatriate workers can now renew their labour permits until they reach the age of 65.

الآن يمكن أن يجدد العمّال المغتربون تصاريح عملهم حتّى يصلون إلى عمر ال65. (SR)

(now can that renew the-workers the-expatriate permits labour until they-reach the-age of 65)

Such simple and basic grammar rules need to be integrated in the systems to have well-formed translations.

Dual and plural forms also need to be adopted, especially with figures and number expressions. A dual that is an object of a preposition should have the case ending بن , and its modifier should have the same case, but the output does not cater to these rules,

e.g. في السنتان القادمتان (SR). In the following example, one rule of compounding is followed in translating the subject, but the very rule is violated in the dual number expression and the noun it modifies:

OPEC delegates said Monday that the group planned to allow its members to provide up to two million barrels a day of additional crude oil if the market needs it.
 قال مندوبو منظمة أوبك يوم الإثنين أن الجماعة خطت للسماح لأعضائها أن يزودوا حتى مليونان برميلاً يوم نפט خام إضافي إذا احتاج السوق له. (SR)
 (said delegates organization OPEC day the-Monday that the-group planned to-allow to-its-members to provide up-to million-DUAL barrels day oil crude additional if need the-market to-it)

Although the output is of a high level of informativeness and grammaticality, the case of both the number expression and the noun it modifies are wrong. Since the expression of number is dual preceded by a preposition, the grammatical rule states that it should end with *ين* and the final *ن* has to be deleted when the expression modifies a noun. The case ending of the noun *برميل* (barrel) is correct if it was another number expression. Agreement between number expressions and nouns they modify has elaborate rules that need to be observed by the systems. The same problem can be seen in the following example:

Government official says rents will come down in next two years.
 مسؤول حكومي يقول سياتي إيجارات إلى أسفل في تالية اثنان سنون. (SN)
 (government official say will-come-MASC-SING rents to down in next-FEM two years)

The system here does not adopt the dual as the correct translation of *two years*, but it arranged the words of the output according to the sequence of the original. The special usage of the plural form of 'year' in Arabic will be discussed below under stylistic issues.

5.3 Style and Spelling

Although style does not have much significance in determining a good MT translation, opting for a stylistically marked translation is unnecessary. It seems that the system has some marked words as established equivalents. Using TL words and expressions that have a literary character would give the output unneeded higher profile:

Ali arrived yesterday.
 علي وصل بالأمس. (SN)

The system uses a literary form constructed of a prepositional phrase of the unmarked form. There is no need here to use this form. Literature is not the domain of MT yet so as to consider style in the output.

All translations by Systran unnecessarily use the marked plural form of 'year'. This form does not even adapt to the number expression next to it:

NASA Planning Return to Moon Within 13 Years.
 الإدارة الوطنية للملاحة الجوية والفضاء تخطط عودة إلى قمر ضمن 13 سنون. (SN)
 (the-management the-national for-the-navigation and the-space planning returning to moon in 13 'years')

Another point in this respect is the translation of journalistic and news reporting style. News stories and reports usually keep the source at the end of the sentence, e.g. *the company said in a release* in the example below. Keeping the same pattern would produce a translation that lacks the flow of the TL, especially with the mistranslation of the key words of the SL text:

Emaar Properties PJSC, the world's largest property developer by market capitalisation, reported a 255 per cent jump in its net profits yesterday to Dh3.69 billion for the first nine months of 2005, compared to Dh1.03 billion for the same period of the previous year, the company said in a release.

بي جيه إس سي خواصّ إمار، أكبر مستثمر عقاريّ للعالم بتمويل السوق، أبلغ عن القفزة بمقدار 255 بالمئة في صافي ربحها أمس إلى 3.69Dh بليون لأول تسع أشهر من ال2005، بالمقارنة ب1.03Dh بليون للفترة النفس للسنة السابقة، الشركة قالت في تحرير. (SR)

(be jeh iss cee characteristics imaar, largest investor property-ADJ for-the-world by-financing the-market, reported about the-jump by-amount 255 per-the-cent in net profits-its yesterday to Dh3.69 billion for-the-first nine months from the-2005, by-comparison with Dh1.03 billion for-the-period the-same for-the-year the-previous, the-company said in release)

The output of the three systems copied the input style in all the five sentences of this kind. An ideal translation would reverse this pattern and place the reporting source first.

Now I will touch on spelling issues whose discussion, I hope, will help in improving the system output. There are two shortcomings in Google's translations regarding the writing format and spelling. First, all forms of the letter *hamza* are not written correctly and thus all words with this letter are misspelled and distorted. Second, the system does not differentiate between Arabic final 't' (taa? marbuta) and final 'h' (haa?). For example, *اداره* is meant to read 'management' while it can be read 'he runs it', and *اخذ مكانه* can mean 'he-took his place' or 'assumed a status'.

In the section about the Arabic/English translation mode, a remark was made regarding systems not being able to differentiate in their output between two letters: the final *y* (yaa? mutatarifa) and the final *alif* (alif maqsura). Google's output in this mode does not at all differentiate in using these letters, producing misspelled words that can affect the meaning negatively e.g. *حتي، الي، دعوي*. It is important to make such differentiation to have the words in this mode correctly spelled and the meaning of the translation unaffected. This has an implication on the meaning of words and consequently on the semantic well-formedness of the translation.

Finally, Sakhr gets the credit for having diacritics in its system output, but its showing the doubling sign (shiddah) in too many places where it is not that necessary, makes the output too crowded.

6. CONCLUSION

One important feature of machine translation is to maximize meaning, so that minimum effort and less time are required to understand the output. The user should not put forth too much effort to link together the different elements of the translation. Furthermore, MT business is in a development phase that just having a gist should not be acceptable. A good MT system should endeavour to go a further step beyond the gist level. Rules need to be developed and refined so that the output can reach the finest product possible with minimum post-editing needed. The less post-editing is required, the more

successful the translation is, and the less time is spent and less work is done to produce the final translation.

Whereas some of the output of the three systems investigated here is found to give full coherent meaning, and in some cases grammatically correct, translations in general have a problem in one or more aspects. Incoherent translations seem to be produced due to some fault or deficiency in one (or more) phase or component of the MT system.

The aspects and issues discussed herein need to have rules and filters developed. The lexical environment and collocations are very important guides that need to be adopted to help deciding the meaning and choosing the right equivalent.

Addition and deletion are problems that Google needs to look at so that its output is a reproduction of the SL text with no elements deleted or extra elements added. Spelling is another issue that requires attention.

Finally, the too literal translation by the Systran system in comparison with the other two systems, so many words remaining untranslated, as well as mistranslation of lexical items, wrong word order and awkward syntax, makes the post-editing needed for Systran's output so comprehensive and time consuming. The output is a distorted language with no cohesion and coherence. Systran translations need to be investigated in a more detailed study to help in identifying and rectifying the large number of errors produced.

NOTES

-
- ¹ Google online MT is a beta service that is already available and has been used for some time to be included in the study.
 - ² Due to limitation on the size and number of texts that can be translated online, I used, with kind permission from Sakhr Software Company, its SET system which the Tarjim website uses in carrying out the translation.
 - ³ The word *الحياة* is transliterated as *Al-Hayat*, which can lead to the conclusion that the system considered it as a proper noun.
 - ⁴ The name and titles in the second set are translated literally, not into their established English counterparts.
 - ⁵ Foreign names are transliterated in Arabic in slightly different forms, and thus the system should have all possible variants.
 - ⁶ Note that the place of insertion makes a difference in the meaning.
 - ⁷ The place of insertion makes a difference in the meaning.
 - ⁸ It means *hence...has/have* when followed by a verb in the past tense, and *hence...might* if followed by a verb in the present tense.
 - ⁹ The Google system has a problem in writing the Arabic letter *hamza*, and its output is thus misspelled and distorted.
 - ¹⁰ Transliterated words are put in square brackets by the system.
 - ¹¹ As diacritics are not shown and no subject is available, the translation of the verb form can also mean you-maintained-SING-FEM, you-maintained-SING-MASC, or she-maintained.
 - ¹² All cities are feminine in Arabic, e.g. London, Dubai, Paris; but not all countries. A few countries are masculine, e.g. Iraq and Lebanon.
 - ¹³ As diacritics are not shown and no subject is available, the Arabic translation of the verb *sit* can be you-SING-FEM-sat, you-SING-MASC-sat, or she-sat
 - ¹⁴ Due to unavailability of subject and diacritics, the Arabic translation of the verb can also mean you-SING-FEM-passed, you-SING-MASC-passed, or she-passed.

REFERENCES

- Al-Munjid Dictionary* (2002). 39th edition. Beirut: Dar Al-Mashreq.
Al-Qamoos Al-Muheet Dictionary (1994). 4th edition. Beirut: Mu'asasat Al-Risalah.

- Al-Salman, Saleh (2004). The Effectiveness of Machine Translation. In *IJAIS, International Journal of Arabic-English Studies*, 5, 145-160.
- Alsharaf, Haitham, S. Cardy, P. Greenfield, and Y. Shen (2004). Problems and Solutions in Machine Translation Involving Arabic, Chinese and French. In *Proceedings of the International Conference on Information Technology: Coding and Computing (ITTC04)*. No page numbers.
- Al-Waseet Dictionary* (1972). 2nd edition. Istanbul: Islamic Library.
- Al-Wasiti, Salman (2005). *al-tafa'ul bayna al-?insaan wa al-?aalah* (interaction between humans and machines). Downloaded from <www.almughtarib.com> on 25 August 2006.
- Chalabi, Achraf (2001). *Sakhr Web-based Arabic/English MT Engine*. Downloaded from <www.elsnet.org/arabic2001/chalabi.pdf> on 25 August 2006.
- Farghaly, Ali and Jean Senellart (2003). Intuitive Coding of the Arabic Lexicon. A paper read at *The MT Summit IX - Workshop: Machine translation for semitic languages, New Orleans, USA, 23 September 2003*. Downloaded from <<http://webpace.isi.edu/mt-archive/languages-1.htm>>
- Gaspari, Federico (2004). Integrating on-line MT services into monolingual web-sites for dissemination purposes: An evaluation perspective'. In *Proceedings of the 9th EAMT Workshop: Broadening horizons of machine translation and its applications*. Valletta, Malta. 62-72.
- Google On-line Translation Page*. <http://www.google.ae/language_tools?hl=en> visited June-August 2006.
- Lisan Al-Arab Dictionary* (1999). 3rd edition. Bierut: Dar Ihya Al-Turath Al-Arabi.
- Sakhr's Online Translation Website, Tarjim*. <www.tarjim.com> visited June-August 2006.
- Systran On-line Translation Website*. <www.systranet.com> visited June-August 2006.