
Arabic Morphological Generation and its Impact on the Quality of Machine Translation to Arabic

Ahmed Guessoum¹ and Rached Zantout²

¹ *Department of Computer Science, The University of Sharjah, P.O. Box 27272, Sharjah, UAE*
guessoum@sharjah.ac.ae

² *Faculty of Engineering, Department of Computer & Communications Eng., Faculty of Engineering, Hariri Canadian University, Mechref, P.O.Box:10 – Damour, Chouf 2010, Lebanon,*
rached@cyberia.net.lb

Abstract: The aim of this chapter is to highlight the complexity and importance of Arabic morphological information in an Arabic Machine Translation (AMT) system, i.e. a system that translates to or from Arabic. We summarize Arabic morphology and introduce the main morphological information that we have found relevant to machine translation to Arabic and categorize it into various types of features. In order to show the impact of these morphological features on machine translation quality, we have adopted an approach whereby we relate each of them to the quality of the translation. This leads us, through a statistical analysis of the test data, to a characterization of which features are more important in terms of their impact on the quality of the translation of a given AMT system (AMTS). The approach has been implemented and applied to evaluating an English-to-Arabic web-based MT system. The results of the evaluation of this system are presented, conclusions are drawn, and recommendations for improving their outputs made

15.1 Introduction

Translating between different languages is a very important discipline. The estimated value of the world market for translation was U.S. \$20 billion according to the Gartner Group (Stamford, CT, USA) with an annual growth rate of 14.6% (Van der Meer, 2003). In 2004, the human translation market was estimated to be \$1 billion (Oren, 2004), while the machine translation market was forecast to be in the \$100 million range. MT software was reported to be responsible for the completion of between 30 and 50 percent of a Machine translation task automatically (ECL, 1996) and (Hedberg, 1995). MT software was also estimated to cut the cost of translation by two thirds.

Evaluation of Natural Language Processing (NLP) systems is currently a field of research on its own. Various researchers have stressed the importance of component-based evaluation and detailed error analyses (Arnold et al., 1993; Hedberg, 1995; Nyberg et al., 1994). Since MT systems combine lexical analyzers, morphological analyzers, parsers, semantic disambiguation modules, generators, and pragmatic analysis modules, it is important to be able to evaluate these various components individually as well as to evaluate the overall system. The main difficulty here is that, in most of the cases, evaluators do not have access to the individual components of the system under evaluation and are therefore forced into black-box evaluation. This means that an error in the output of the system cannot be attributed to one of the components since it can be due to one or many errors in one or more of the components of an NLP system.

In (Van Slype, 1979) evaluation is subdivided into two main categories: macro-evaluation and micro-evaluation. Among the macro-evaluation assessment components that are affected by morphological errors are the cognitive level components such as intelligibility, fidelity, coherence, usability, and acceptability of a translation. Among the micro-evaluation methods affected by morphological errors are the grammatical symptomatic components such as the analysis of grammatical errors found in the target output. (Chaumier et al., 1977) suggest an even finer scrutiny of the grammatical (sub-) constructs in the source and target texts, e.g., noun phrases, adjectival and verb phrases, object complements, adverbial complements, etc.

In general, there seems to be an agreement as to the following aspects that should be evaluated in any MT system: *adequacy*, which is the extent to which the meaning of the source text is rendered in the translated text; *fluency*, which is the extent to which the target text appeals to a native speaker in terms of well-formedness of the target text, grammatical correctness, absence of misspellings, adherence to common language usage of terms, and meaningfulness within the context (White et al., 1994); *informativeness*, which assesses the extent to which the translated text conveys enough information from the source text as to enable evaluators to answer various questions about the latter based on the translated text; and *intelligibility* (Arnold et al., 1993), which is strongly related to informativeness, though directly affected by grammatical errors and mistranslated or missing words. In (Nyberg et al., 1994) the authors from the KANT (Nyberg et al., 1992) team introduce a methodology based on evaluation metrics for knowledge-based MT. The evaluation criteria they consider are: *completeness*, which measures the ability of a system to produce an output for every input; *correctness*, which measures the ability of a system to produce a correct output for every input; and *stylistics*, which measures the appropriateness of the lexical, grammatical, and other choices made during the translation process. Based on the completeness, correctness, and stylistics criteria, the authors then defined four evaluation criteria, which test, as percentages, the Analysis Coverage, Analysis Correctness, Generation Coverage, and Generation Correctness. These four percentages then get multiplied,

yielding the *Translation Correctness*, which measures the overall quality of the system. (Guessoum & Zantout, 2001) and (Guessoum & Zantout, 2005) introduce a semi-automatic methodology for component evaluation of Arabic MT systems (AMTSS) using a black-box approach. The methodology tests the correctness of each component of an MT system by analyzing carefully selected sentences translated by an MT system. Weighted averages are then computed and scores are derived for each component of a system under evaluation. An overall score for the system is also calculated. The weighted averages were shown to be indicators of what components of the system are the faultiest and therefore would need immediate attention by the developers. The difficulty in this approach is the ability to come up with a large number of test sentences that would test each component of the MT system.

The aim of this chapter is to highlight the complexity and importance of Arabic morphological information in an Arabic Machine Translation (AMT) system. In Section 15.2 we summarize Arabic morphology and in Section 15.3 we introduce the main morphological information that we have found relevant to machine translation to Arabic. We show how various aspects of Arabic morphology reflect important lexical, syntactic, semantic, and pragmatic aspects of a sentence in translation. We also categorize this information into various types of features. In Section 15.4 we show how, through a statistical analysis of a bilingual corpus consisting of English source text and machine-produced Arabic target text, we could suggest a characterization of which morphological features are more important in terms of their impact on the quality of the translation of a given AMT system (AMTS). In Section 15.5 conclusions are drawn, and recommendations for improving the outputs of AMT systems made.

15.2 Basics of Arabic Morphology

The information in this section was derived through readings in (Dahdah, 1995), (Hamalawy, 1996), and (Rajhi, 1979). Arabic words are grouped into three main categories: Nouns (الأسماء), Verbs (الأفعال) and Prepositions (الأحرف). The Noun and Verb categories consist of subcategories that affect how the word is used in the sentence and how it changes within the context of the sentence and the other words in the same sentence. Each subcategory obeys certain rules of morphology that detail whether a word can be used in a certain context and how its form changes in that context. The difference between words in subcategories can be as subtle as the presence (or absence) of a vowel or as clear as the addition (or removal) of letters to the word when it moves from one subcategory to another. Although the Preposition category contains subcategories, prepositions in Arabic do not, in general, change forms.

Morphology for Arabic is a tool that enables the language to grow and develop. Morphology, in general, is defined as producing a word from another by changing

it so that it fits a certain new meaning. In Arabic, morphology is divided into four categories of derivations, the small (الإشتقاق الأصغر), the large (الإشتقاق الكبير), the larger (الإشتقاق الأكبر) and the largest (الإشتقاق الكبار). The small morphology produces one word from another but keeps similarities between the two words in their pronunciation and meaning (e.g. علم¹ (çilm, science)→ عالم (çAlim, scientist)). The large morphology produces a word from another by exchanging the letters in the roots of the words (علم (çilm, science)→ عمل (çamal, work)). The larger morphology produces a word from another by changing one (or more letters) and keeping the same meaning (عنوان (çnwAn, address)→ عنوان (çulwAn, address)). The largest morphology produces a word from a group of words such as the contraction بسملة (basmalh) from بسم الله الرحمن الرحيم (bismi All~Ah Alr~aHmAn Ar~aHym), “In the name of Allah, The Compassionate, The Merciful”). By far, the mostly used type of morphology in Arabic is the small morphology.

Small morphology can act on a Noun or a Verb. Any Noun or Verb in Arabic consists of a root and added letters. The roots for Arabic words have traditionally been considered to consist of three letters (the mostly used type of roots in Arabic) or four letters. Like English and French, in Arabic, letters can be added to the beginning and/or to the end of the root. However, unlike English and French, in Arabic, letters can be added inside (between the letters of) a root. This is one of the complexities that make Arabic a harder language to analyze or generate morphologically.

Fortunately, for computational linguists interested in developing Arabic language tools on computers, Arabic is a structured language. Basically, verbs and nouns cannot accept additions of all letters in the Arabic language at all places inside, at the beginning or the end of a root. In this chapter we will explain some of the rules pertaining to verbs in the Arabic language. The reader should bear in mind that Nouns obey similar rules. The reader should also bear in mind that the rules described below will not be exhaustive even for Arabic verbs as the purpose behind the explanation is to give the reader an appreciation of the complexity of Arabic morphology rather than to enumerate all the rules governing Arabic morphology.

In Arabic, there are certain letters that can come at the beginning of the root (prefixes); these letters are grouped in the Arabic word سألنتي (s, A, l, t, n, y). The letters that can be added to the end of the root (suffixes) are grouped in the Arabic word أوهمتتي (A, w, h, m, t, n, y). There is an upper limit on the number of prefixes (four letters) that can be added to a root. In some cases, a letter can be repeated as a prefix or a suffix. Letters that can be added to the inside of a root (infixes) have a more complicated set of rules. First, a group of letters ا (A), و (w), and ي (y)

¹ In the rest of the paper, any Arabic sentence will be followed by its transliteration using the scheme followed throughout the book .

(called *أحرف العلة*), while being part of the root, can disappear from an Arabic verb if the verb is in the imperative form. Second, only one or two letters can be added inside the root. Third, infixes are grouped in the Arabic word *اتوني* (A, t, w, n, y). Fourth, certain verb forms can be produced by repeating the same letter of the root.

Arabic morphology is a very structured process. For example, a verb can undergo morphology based on moulds that take any root and transform it into the corresponding verb by adding prefixes, suffixes or infixes in order to convey the meaning of the verb. For example, in order to specify that more than two people are writing to each other, the root *كتب* (*kataba*, (he) wrote) is used. The suffix *ون* (*wn*) is added to it to indicate that the verb is being done by more than two people²; the prefix *ي* (*y*) is added to indicate that the verb is in the present tense; and the infix *ا* (*A*) is added to indicate that they are writing to each other. Thus the verb obtained is *يكتابون* (*yukAtibwn*). If the same meaning is to be conveyed but, now, instead of the group of people writing to each other we want to say that they play with each other, it is only necessary to replace the letters of the verb root *write* *كتب* (*kataba*) with those of the verb root *play* *لعب* (*laçiba*) to obtain the verb *يلعبون* (*yulAçibwn*). In a similar manner, if the group consisted of two instead of more than two members then the only change needed is to use the suffix *ان* (*An*) instead of the suffix *ون* (*wn*). The different forms that can be used with a root to produce an Arabic verb have been classified differently in the literature. One such classification (Fowzan et al., 2000) enumerates 129 Morphological Patterns which can be used to generate verbs from roots.

In Arabic, the Noun or Verb will have different forms if the subject or object is masculine or feminine. Also differences in the forms can exist if the sentence refers to one person or a group of two or a group of more than two.

15.3 Arabic Morphological Generation as a Repository of Morphological, Syntactic, Semantic, and Pragmatic Information (in an AMTS)

It is well-known that Machine Translation is a complex process. It is ideally the result of analyzing the source text morphologically, lexically, syntactically, semantically — and even pragmatically and stylistically if needed, and producing its equivalent target text using all of these linguistic dimensions. The target language and the complexity of its morphology and grammar can make this machine translation process even more complex. This is indeed the case for Arabic, where, due to the complexity of the morphology, the generation of correct Arabic words must

² Recall that Arabic has two forms for the plural: the dual and the non-dual plural (more than two people).

take into account and reproduce all the linguistic information acquired from the analysis of the source text, be the source language English or any other.³

Consider for instance a simple sentence like “The girls wrote the beautiful essays”. To translate it to Arabic, the morphological generator needs to

1. add the prefix ال (Al, the) to the word بنات (banAt, girls) which itself is the result of adding the infix ا (A) to obtain the plural of بنت (bint, girl);
2. add the suffix ت (t) to the basic verb form كتب (kataba, he wrote) to produce كتبت (katabat, she wrote) for past tense, feminine form;
3. add the prefix ال (Al, the) and suffix ات (At, feminine plural) to the word مقال (maqAl, an article/essay) because, in Arabic, the masculine word for article/essay has a feminine plural form;
4. add the prefix ال (Al, the) and suffix “ة” (h, for feminine) to the adjective جميل (jamyil, beautiful (masculine singular)) to obtain the feminine form of the adjective, for gender concordance with the plural word for articles/essays.

The generated Arabic sentence would therefore be

كتبت البنات المقالات الجميلة.

katabat AlbnAtu AlmqAlAt Aljmylh

Wrote (fem.) the-girls (fem. plural) the-articles (fem. plural) the-beautiful (fem.).

In the above example, the verb and the subject have to match in gender; the noun and the adjective match in gender but also with respect to the definite case. It is clear that the morphological generator needs to take into account lexical and syntactic information in addition to the fact that a word re-ordering needs to be introduced by the “transfer” module.⁴ As such, by reading the Arabic sentence, we can immediately tell (i.e. from the output of the morphological generator) whether the words are lexically and syntactically correct. In fact, even the meaning can be affected, as will be explained in the coming sections, if the morphological generator does not receive this information from other modules (such as the parser) in the machine translation system or does not correctly reproduce it.

In more complex examples, pragmatic knowledge could be used, such as reference resolution, so that the proper form of a word is generated. Consider, for example, the following sentence, its translation, and gloss:

This is your room - it's rather small هذه هي حجرتك - هي صغيرة إلى حد ما

hađihi hiya Hujratuka – hiya Saýyraħũ Āily Had~ĩ mA

This it (is⁵) your-room - it (feminine) (is) small (feminine) to limit some

³ Despite the fact that the discussion in this chapter is about Arabic morphological generation, independently of the source language in any machine translation process, all the examples and the implementation will be about machine translation from English to Arabic.

⁴ We call it “transfer” module no matter what actual approach is adopted in the machine translation system.

⁵ The auxiliary “is” is implicit in Arabic.

In this case, the demonstrative pronoun “*this*” should be translated as هذه (haḏihi, (feminine) *this*) and not هذا (haḏA, (masculine) *this*) since حجرة (Hujrah, room) is feminine. The gender needs also to be conveyed in the second هي (hiya, (feminine) *it*) and صغيرة (Sayyrah, (feminine) *small*). Again, an error in the resolution of the references would be confusing or misleading. For instance, if the previous source sentence gets incorrectly translated as

(*) هذه هي حجرتك – هو صغير إلى حد ما

haḏihi hiya Hujratuka – huwa Sayyrū Āily Had-ī mA

This it (is) your-room - it (masc.) (is) small (masc.) to limit some

This translation would convey a completely different meaning. Indeed, the reader would believe that the speaker mentions the room of the listener but goes on/back to talking about some other person describing *him* as small to some extent. If a male person happens to be mentioned in the context of the sentence, the confusion would become complete!

In the context of a black-box evaluation, it is *not* possible to find out the faulty component(s) of a machine translation system that produces an output like the one in the last example. We cannot be sure whether the morphological generator received all the needed information to produce the correct words, and hence it would be the faulty component, or if it did not receive enough information from the other components in the MT system (parser, transfer, etc.) as to generate the correct words. In all cases, what we argue is that the presence or absence of morphological information can affect quite seriously the quality of an MT system.

From an analysis of a large number of sentences, as will be explained in the next section, we have singled out and categorized various types of morphological features that are important in Machine Translation and which can affect its quality. In fact, this is exactly the criterion for singling them out: we selected a feature if its improper handling affects the sentence quality.

These features are now presented, with clarifying examples, and will be further analyzed in the coming sections on an actual Arabic MT system.

1. Definite / Indefinite Nouns (A)

As explained earlier, an indefinite Arabic noun can be made definite by adding the article ال to it. From our analysis of the outputs of various AMT systems, this is quite frequently not done correctly. The result can be objectionable or even unclear.

| | |
|-----------------------------------|------------------------------------|
| Monkeys are very agile climbers. | القرود متسلقون رشيقون جدًا. |
| ... afflicts women more than men. | ... يبتلي النساء أكثر من رجال. (*) |

In the first example the prefix definite article ال (Al) is correctly added to the noun قرود (quruwd, monkeys). In the second example, both words for men and women

should be definite in Arabic. The translation correctly places the definite article for “(the) women” (النساء , Aln~isA') but not for “men” (رجال , rijAl).

2. Case Ending (B)

One of the complexities of Arabic grammar is that words get inflected depending on the case (nominative, accusative, or genitive), the number (singular, dual, or plural), and the gender (masculine or feminine). The generation of the correct form of the word depending on these features must obviously be done very carefully and is in fact a common mistake among native speakers nowadays! The improper handling of the case ending results in unpleasant sentences and sometimes it modifies the sentence meaning entirely, especially when the word order is changed.⁶

| | |
|--|---|
| ... the massive aerial bombardment of military targets continued unabated. | القصف الجويّ الضخم للأهداف العسكريّة استمرّ قويّاً. |
| He abided in the wilderness for forty days. | أقام في الصحراء لمدة أربعون يوم. (*) |

In the first sentence, the Arabic adjective قويّاً (qawiy~Aā , massive/strong) correctly appears in the accusative form. However, the numeral أربعون (Ārbaʿawn, forty) should be in the genitive form while it appears in the nominative.

3. Imperative Mood (C)

A common error found in Arabic MT systems is the incorrect translation of verbs in the imperative mood into verbs in the present tense of the indicative mode, often with the wrong pronoun (he) being used. This is the case with the system we have evaluated for the purposes of this work.

| | |
|-------------------|------------------|
| Please contact... | (*) من فضلك يتصل |
|-------------------|------------------|

Here the verb contact is incorrectly translated as يتصل (yat~aSil, he contacts). It should rather be the imperative form of the verb (i.e., Āt~aSil, contact).

4. Verb Tenses (D)

This is another error commonly found in AMT systems. It is often coupled with the incorrect use of the pronoun (he). Obviously, enough morphological information needs to be made available to the morphological generator not to fall into this trap. Sometimes, this mistake shows up when the present and past tenses have the same form for a given verb in the source language (English in our case). However, the error would probably reflect an improper syntactic parsing of the source sentence.

⁶ Word order modification is a fairly common thing to do in Arabic; it is usually used to give a different emphasis in the sentence.

| | |
|--------------------------------------|-----------------------------------|
| Why don't we put the bed ... | لماذا لا نضع السرير ... |
| The anti-war agitation has begun ... | (*) الثورة المعارضة للحرب تبدأ... |

In the first example, “we put” is correctly translated as نضع (naDaçu, we put) in the present tense whereas, in the second example, “has begun” somehow gets translated to تبدأ (tabdaÂu, (it) begins).

5. Expressions (E)

Handling common expressions often requires the application of prepositions, the definite article, pronouns, etc., to various categories of words. If this is not carefully done, the result will be incorrect words (in the context) or even morphologically ill-constructed words as explained in item 8 below. Of course, most of the time a word-to-word translation of these expressions gives appalling results morphologically, syntactically, and semantically.

| | |
|--------------------------------|--------------------------|
| For a man of 80... | بالنسبة لرجل في الـ 80 |
| ... in the 25 to 40 age group. | (*) في 25 إلى 40 جيل ... |

The first sentence is correctly translated with the proper Arabic preposition في (fy, in). The second sentence is badly translated, which results in the meaning “in 25 to 40 generations”!

6. Pronouns (F)

Another problem is the proper handling of pronouns. Pronouns can appear either suffixed to a word or separated from it, based on various morphological and syntactic rules. The pronouns may take one of several forms depending on the features mentioned earlier, namely the case, gender, and number.

6.1. Pronoun-Related Concordance (Case Ending)

The morphology of a word can get affected by a pronoun in a sentence. For instance, in the first example below, “as an afterthought” is translated as مستدركة (mustadrikaĥā, as an afterthought of hers) the last letter of this word reflecting the fact that the subject is feminine. If the subject was a plural one like “They”, the word مستدركة would become مستدركات (mustadrikAt, for the feminine plural) or مستدركين (mustadrikyn, for the non-feminine plural). Similar changes would occur if the subject was dual (two people), etc. In the second example, طافيا (TAfiyAā, afloat) appears incorrectly in the masculine singular form; it should be طافية (TAfiyahā, afloat (feminine singular)).

All such morphological information needs to be available and generated for the sentence to be correct.

| | |
|---|--------------------------------------|
| She only asked me ... as an afterthought. | دعتني ... مستدركة . |
| She spent seven days afloat on a raft. | (*) قضت سبعة أياماً طافياً على قارب. |

6.2. Gender Concordance (and Pronoun Resolution)

Also a common source of error is gender concordance, where a pronoun needs to reflect the gender, or other more general aspects of pronoun resolution such as number. In the example below, حجرتك (Hujratuk, your room) is feminine but is incorrectly translated using the pronoun هو (huwa, it (masculine)) and صغير (Sayyr, small (masculine)).

| | |
|--|--|
| This is your room - it's rather small... | (* هذه هي حجرتك - هو صغير إلى حد ما... |
|--|--|

6.3. Unnecessary Generation of Pronouns

From our experience with AMT systems, this error is quite a common one. A pronoun is very frequently generated when it is not needed. Indeed, in Arabic a verb implicitly indicates the person (singular, dual or plural; 1st, 2nd, or 3rd; etc.). As such the addition of a pronoun before the verb is often a redundancy (unless there is an emphasis to be conveyed) and sounds heavy. In the example below, سَيُجْمَعُونَ (sayujmaçwun, (they) will be reunited) has the suffix ون which indicates that the subject is they (non-feminine plural taken by default here). Adding the pronoun to such a verb would convey a meaning like “It is they who will be reunited in the afterlife”, which in the case of this particular sentence and its religious connotation, is probably quite unacceptable.

| | |
|---------------------------------------|-------------------------------------|
| They'll be reunited in the afterlife. | (هم) سَيُجْمَعُونَ فِي الْآخِرَةِ . |
|---------------------------------------|-------------------------------------|

6.4. Incorrect Pronoun

Sometimes, the pronoun is not obvious to guess from the sentence; it is understood from the context or by commonsense. In the example below, “It’s advisable to book seats...” would probably mean “It’s advisable for you to book seats...” or “It’s advisable for one to book his/her seats...”. A common mistake in AMT systems is to simply translate the verb “to book” as يحجز (yaHjizu, he books) conveying the following meaning for the sentence “It’s advisable for him to book his seats...”. This level of morphological information detail is indeed needed to render the semantics of the sentence. A human translator would most probably translate the sentence into:

من المستحسن حجز المقاعد على الأقل أسبوع مقدماً

| | |
|--|--|
| It's advisable to book seats at least a week in advance. | (* هو مستحسن أن يحجز المقاعد على الأقل أسبوع مقدماً. |
|--|--|

7. Number Concordance (G)

This feature should be clear from the above explanations. Nouns, verbs, adjectives, and pronouns match with respect to number. In the examples below, لحضارتين (liHaDAratayni, of two civilizations, genitive form) is in the dual form which is different from لحضارة (liHaDArah, of one civilization). As such, the adjective must be in the dual genitive form عظيمنتين (çDymatayni, (two) great). In the

second example, متلهف (mutalah~if, agog) incorrectly appears in the (default) masculine singular form, which would reflect that the subject is “he” (instead of we).

| | |
|---------------------------------|-----------------------------|
| ... of two great civilizations. | ... لحضارتين عظيمتين. |
| We waited agog for news. | (*) انتظرنا متلهف للأخبار . |

8. Constructed Words (H)

In a number of cases that we have come across, ill-constructed words were generated. This clearly reflects errors that are intrinsic to the morphological generator. For instance, the output in the first example below, an invalid word كبعيدًا (kabaçydAã, as far afield) is generated. This word has probably been constructed by adding the prefix ك (k, as) to the word بعيدًا (baçydAã, far) giving a form which is not correct in Arabic. A similar process was followed in the second example where the suffix ني (ny, me) instead of ي (y, me/my) was incorrectly combined with the word إغضاب (AγDAb, aggravating).

| | |
|--------------------------------|--------------------------|
| ... as far afield as Japan ... | ... كبعيدًا كاليابان ... |
| Stop aggravating me... | توقف عن إغضابني... |

9. Inadequate Prepositions (I)

A preposition can prefix a word (e.g. prepositions ل , ك , ب) or can be separated from it (e.g. prepositions من , عن , على). If one is not careful, prepositions may be incorrectly translated. For example, with the AMT system we have evaluated, “at” in the first example below, was translated as في (fy, in) instead of ب (bi, with/at). Likewise, the incorrect preposition ل (li, for) was selected instead of عن (çn, of) to prefix the word حقوق (Huquwq, rights).

| | |
|---|---------------------------------------|
| ... at affordable prices. | (*) ... في الأسعار المتاحة. |
| She is renowned for her advocacy of human rights. | (*) هي مشهورة بدفاعها لحقوق الإنسان . |

10. Use of the Improper Grammatical Category (J)

This type of error, as we will see below, does occur fairly commonly with AMT systems. It confirms what we have concluded in previous work (Guessoum & Zantout, 2001) and (Guessoum & Zantout, 2005) that the AMT systems we have evaluated follow an improved form of direct MT, although some of them claim that they use transfer-based MT. In the first example below, the AMT system we have evaluated, translates “forty-three” textually as ثلاثة وأربعون (0alA0ahū wa Ârbaçwn) instead of producing the correct form في الثالثة و الأربعين (fy Al0Ali0ahi wa AlÂrbaçyn, in the forty third (year)).

| | |
|--|---|
| ... and at forty-three, somehow ageless. | (*) ... وفي ثلاثة و أربعون من عمرها، بطريقة ما دائمة الشباب. |
| She asked the question expecting an affirmative. | (*) سألت السؤال تنتظر موافقة . |

The above features are what we have singled out as features to be looked at when evaluating the Arabic morphological generation module of an AMT system. The approach adopted in our work is now presented.

15.4 Analysis of Arabic Morphological Generation Features in an Arabic MT System

In order to study the impact of Arabic morphological generation features on the quality of MT to Arabic in an AMTS, we have collected in Phase 1 English sentences by looking up 1056 English words online using the Cambridge Dictionaries site at <http://dictionary.cambridge.org>. Out of the 1056 words, 781 were found to have sample sentences in the online dictionary. Out of these, 756 could be translated to Arabic using the web-based AMTS *Ajeeb* (<http://www.ajeep.com>). These Arabic sentences have then gone through Phase 2. In Phase 2, we analyzed all the pairs of English and Arabic sentences looking for the various types of morphological features that we could single out as important in AMT. These are the various types of features that were presented in Section 15.3. Having a classification of the morphological information that is relevant to AMT, we needed to see how frequent the types are and how much they affected the quality of the output of a given AMTS.

One aspect we have mentioned earlier is that the errors found in the translation may be due to the morphological generator, syntactic parser, or any other module of the AMT system. However, what is relevant to our work is that whatever the source of the error, it is reflected at the morphological generation level as explained in the previous section. As we wanted to find a correlation between the type of the error and the quality of the translation, we had to be very careful in our evaluation approach. In fact, we have followed a number of steps.

First, we kept only the sentences or sentence chunks which contained at least one error related to the morphological features mentioned in Section 15.3. Errors like wrong word order have not been considered as they are purely syntactic (and therefore most probably not related to the morphological generator).

In the second step, we discarded all the sentences that contained more than two errors of the morphological types of interest. This is to avoid confusion as to which type affects the quality of the translation most, and for how much. We ended up with 177 sentences that contained one or two errors related to the categorized morphological features. We then tagged each sentence with A, B, or ..., J, where A stands for the feature type "Definite/Indefinite", B for "Case Ending", etc.,

up to J for “Use of Improper Grammatical Category”, as defined in Section 15.3. As some of the sentences may have two errors, not just one, we decided to assign only one tag/letter considering that error which most affects the sentence and, for simplicity, (mentally) correcting the other one.⁷ As a result, each sentence received exactly one tag.

At this point we were ready for evaluating the selected pairs of sentences for adequacy and correctness of the machine translation. This has been done by human experts who were asked to assign a value between 0 and 5, where 0 means completely unacceptable and 5 perfectly clear at first reading while being faithful to the source sentence and sounding correct.

Once a translation quality value was assigned to each tagged sentence, we computed statistics giving the number of sentences afflicted with each type of error (having been assigned a specific tag) as well as the average quality measure (value between 0 and 5, inclusive) for each one of these types. This measure tells us how much an error for that morphological feature affects the quality of the translation. The closer the value to 0 for a particular type of morphological information, the more serious an impact the type has on the quality of the translation. Obviously, the closer the value to 5, the less impact the type has on the quality of the translation.

15.5 Results and Data Analysis

Table 15.1 shows the results of the analysis of the 177 pairs of (tagged) sentences translated using the AMT system mentioned in Section 15.4.

The frequencies of the sentences of types A, B, ...to J, are given in the second row. We clearly see that the largest number of errors is about the handling of pronouns (20.34% of the cases), followed by using improper grammatical category (16.95% of the cases), and then by using inadequate prepositions (15.82% of the

Table 15.1. Quantitative breakdown of morphological errors affecting AMT system quality

| Type of Error | A | B | C | D | E | F | G | H | I | J |
|--------------------|------|-------|-----|------|-----|-------|------|------|-------|-------|
| Number of Cases | 20 | 22 | 7 | 9 | 8 | 36 | 8 | 9 | 28 | 30 |
| % of cases | 11.3 | 12.43 | 3.9 | 5 | 4.5 | 20.34 | 4.5 | 5 | 15.82 | 16.95 |
| Average (Out of 5) | 4.25 | 4.41 | 1 | 2.56 | 2.5 | 2.9 | 3.12 | 2.22 | 2.61 | 1.43 |

⁷ Note that we could refine our evaluation by considering combinations of errors as further affecting the meaning/intelligibility of the target sentence.

cases). On the other hand, the least frequent error type is the handling of the imperative mood (3.9% of the cases).

The frequencies just presented are not meaningful enough on their own if we do not know how serious each of the error types is and how much it affects the quality of the translation. This is what the last row of Table 15.1 tells us. In particular, it shows that the type of morphological error which most affects the quality of the translation (in the case of the AMTS under evaluation) is the handling of the imperative mood. The results tell us that despite the fact that this error occurs in only 3.9% of the cases, whenever it occurs, it drastically affects the meaning of the sentence with an average score of 1 out of 5. The rest of the table can be read in the same way. In particular, only two types of errors are “mild” enough as to produce sentences that are still reasonably adequately translated, with a quality score of more than 4 out of 5. These types of errors are the incorrect handling of Definite/Indefinite nouns and the incorrect case endings. Both of these errors are fairly frequent (>11% of the cases for each one) but do not seriously affect the meaning and correctness of the target sentence.

Table 15.1 is concise enough and, in our opinion, quite useful for a better evaluation of an AMT system and, hopefully, for its improvement. Indeed, the averages we have computed also tell us how much an improvement of the quality of the sentence we would get if we correct that particular type of morphological information error. For instance, correcting the error for the case of the handling of the imperative mood should make the sentence noticeably more comprehensible.

15.6 Conclusion and Recommendations

In this chapter, the importance of Morphological Generation to the clarity of the output of an MT system was emphasized. The complexity of morphology in Arabic was presented through the description of some of the rules that govern Arabic morphology. An analysis of the common types of errors related to Arabic morphological generation was then made and several important types of errors were detailed. An existing commercial AMTS was then evaluated to determine which type(s) of error affected the translation to Arabic using that AMTS. The approach used in this chapter for evaluation identified several types of errors as affecting the output of the AMTS most drastically.

It is expected that the developers of the AMTS under evaluation would benefit from this evaluation by concentrating their research on treating the most common errors and those which affect the output of the AMTS most drastically. The categories of errors that were identified in this chapter can also help developers of AMT systems look for such errors in their output and treat them inside the AMTS. This will lead to a better AMT system that will produce outputs of better quality.

The evaluation of one AMT system in this contribution has revealed important information about output errors and their types. It is recommended that other AMT systems be evaluated in the same manner. This will lead to determining

whether the types of errors that were identified in this chapter are indeed general for many AMT systems and therefore being general features of Arabic to be paid attention to in machine translation to Arabic.

As we happen to have collected a much larger corpus of pairs of translated sentences, we intend to do the evaluation for a much larger part of this corpus so as to reach statistically more conclusive results. Research could also be done on how to automate the evaluation above by using language tools that would be able to analyze the Arabic sentences and identify the types of errors automatically. For example, using an Arabic morphological analyzer, the output words could be automatically analyzed into a set of roots and associated morphological information. Then, a module could be developed that would check the types of errors by analyzing the information for all words in a sentence. Automating parts of the evaluation would allow the treatment of large corpora in shorter times.

Acknowledgements

We would like to thank M. A. Hussein (BSc student at the University of Sharjah) for having collected the sample English sentences from the web site <http://dictionary.cambridge.org> as well as their translations to Arabic using the web-based Machine Translation system Ajeeb (<http://www.ajeab.com>).

References

- Arnold, D.J., Sadler, L.G. & Humphreys, R.L. (1993). Evaluation: An Assessment. In *Machine Translation. Special Issue on Evaluation*, 8(1–2), 1–24.
- Chamnier, J., Mallen, M.C. & Van Slype, G. (1977). Evaluation du Système de Traduction Automatique SYSTRAN; Evaluation de la Qualité de la Traduction. *CEC Report Nr. 4*. Luxembourg.
- Dahdah, A. (1995). معجم تصريف الأفعال العربية [Dictionary of the Conjugation of Arabic Verbs]. Beirut, Lebanon: Librairie du Liban.
- ECL (Equipe Consortium Limited) (1996). *Survey of Machine Translation: Products and Services*. Summary of a report to the European Commission. Retrieved from <http://www2.echo.lu/langeng/rep/mtsurvey/mtsurvey.html>
- Fowzan, M., Al-Harbi, S., Kazdar, S. & Al-Qahtani, H. (2000). المحلل الصرفي للأفعال [Morphological Analysis of Verbs]. B.Sc. project report, King Saoud University, College of Computer and Information Sciences, Computer Sciences Department, Riyadh, Saudi Arabia.
- Guessoum, A. & Zantout, R. (2001). A Methodology for a Semi-Automatic Evaluation of the Language Coverage of Machine Translation System Lexicons. *Machine Translation*, 16(2), 127–149.
- Guessoum, A. & Zantout, R. (2005). A Methodology for Evaluating Arabic Machine Translation Systems. *Machine Translation*, 18(4), 299–335.
- Hamalawy, A. (1996). شذا العرف في فن الصرف [The Art of Morphology]. Beirut, Lebanon: مؤسسة الكتب الثقافية.
- Hedberg, S. (1995). Machine Translation Comes of Age. *Computer Select*, September.

- Van der Meer, J. (2003). At Last Translation Automation Becomes a Reality: An Anthology of the Translation Market. In *Proceedings of International Workshop of the European Association for Machine Translation/Controlled Language Applications Workshop* (pp. 180–184), Dublin, Ireland.
- Nyberg, E. H. III and Mitamura, T. (1994). The KANT System: Fast, Accurate, High-Quality Translation in Practical Domains. In *Proceedings of the 14th International Conference on Computational Linguistics, COLING-92*, pp. 1069-1073.
- Nyberg, E. H. III., Mitamura, T. & Carbonell, J.G. (1994). Evaluation Metrics for Knowledge-Based Machine Translation. In *Proceedings of COLING -94*.
- Oren, T. (December 2004). *Machine Translation and the Global Blogosphere*. Retrieved from <http://windsofchange.net>.
- Rajhi, A. (1979). *دار النهضة العربية للطباعة والنشر* [Applied Morphology]. Lebanon: دار النهضة العربية للطباعة والنشر.
- Van Slype, G. (1979). *Critical Study of Methods for Evaluating the Quality of Machine Translation* (Final Report). Prepared for the Commission of the European Communities, Directorate for General Scientific and Technical Information Management (DG XIII), BR 19142, Brussels.
- White, J., O'Connell, T. & O'Mara, F. (1994). *Machine Translation Program: 3Q94 Evaluation..* In *Proceedings of the November 1994 Meeting of the Advanced Research Projects Agency*. Retrieved from http://ursula.georgetown.edu/mt_web/3Q94FR.htm