A Linguistic Analysis of the Problems of Facebook (see translation) and Google (translate) Applications: A Selected Sample of Corpora

Dr. Heba Abdelraheim Ibrahim Alkady
Lecturer in Linguistics & Translation Studies
Faculty of Arts - South Valley University
A Linguistic Analysis of the Problems of Facebook (see translation) and Google (translate) Applications: A Selected Sample of Corpora

Abstract

Machine Translation or (MT) is considered one of the recent innovations in technology in the field of translation studies (TS). This paper is an attempt to redefine nowadays users’ methods of translating corpora from Arabic into English and vice versa. MT is a computational activity in which the translating process is done by using a bilingual or multilingual data set. Corpus analysis is one of the fastest-growing methodologies in contemporary linguistics. Many institutions and individuals use MT to translate corpora. They resort to computational translation applications offered by Facebook (see translation) and Google (translate) to render their corpora (words, utterances, statements, speech, texts...etc. Regardless of the fatal mistakes sometimes found in the results of MT (For example, those mistakes related to diacritical markers), no one can deny that MT systems are built on gigantic lexical banks and dictionaries. This paper presents a linguistic analysis of the problems of MT namely Facebook (see translation) and Google (translate) applied to a limited sample of corpora.

Keywords:
Machine Translation, Diacritical Markers, Facebook (see translation), Google (translate), Corpora.
دراسة لغوية تحليلية لمشكلات الترجمة الآلية لتطبيقي فيسبوك وجوجل لعينة من المتون

ملخص البحث

تعتبر الترجمة الآلية إحدى صور الإبتكار في الصناعات اللغوية الحاسوبية وتمثل تقنية مذهلة في مجال دراسات وعلم الترجمة. يمثل هذا البحث محاولة التعرف أكثر على وسائل وتطبيقات مستحدثة شائعة بين مستخدمي تطبيقات الترجمة العربية والإنجليزية منظورًا على جديد وتعدد المشكلات التي تواجه هذا النمط الآلي عن طريق تحليل لغوي لعينة من "المتون" المفردات والعبارات والمنشورات الفيسبوكية التي تمثل ظواهر وصعوبات لغوية متنوعة من العربية إلى الإنجليزية والعكس. تتم عملية الترجمة الآلية عن طريق نظام معالجة ثنائي اللغة ورغم نقائص وعيوب الترجمة الآلية التي تعود للسياق اللغوي والخصائص التركيبية النحوية والصرفية والاشتقاقية والدلائية للمفردات التي تمثل صعوبة تطبيقات الترجمة الآلية اعلاميات التشكيك إلا أنها مرجع ووسيلة سريعة يستخدمها كثير من الأفراد والمؤسسات في ترجمة المفردات والنصوص من وإلى الإنجليزية حيث لوحظ ازدياد إقبال المستخدمين على تطبيقات الترجمة الآلية مثل فيسبوك وجوجل.

تهدف الدراسة إلى تقديم تحليل لغوي لبعض المشكلات الشائعة التي تواجه مستخدمي تطبيقي فيسبوك وجوجل تحديداً ولا تعالجها هذه التطبيقات.

الكلمات الدالة:

الترجمة الآلية. علامات التشكيل. تطبيق ترجمة فيسبوك. تطبيق ترجمة جوجل. المتون.
A Linguistic Analysis of the Problems of Facebook (see translation) and Google (translate) Applications:
A Selected Sample of Corpora

Introduction

Arabic Language is spoken by a great ratio of the world’s population. It is a powerful means of social control. It is used as a means of communication among Arabs and non-Arabs as well. Accordingly, Arabic is becoming a crucial language on the Internet due to the increasing number of Arabic speaking online users (Facebookers & Googlers) seeking Arabic content and translation applications, so it cannot be sidelined.

The rapid advance of online services to satisfy users is important to enhance the Internet usage. Google (translate) and Facebook (see translation) applications are examples of this advancement. These applications involve the capability to support multiple languages and provide tools to offer and use multi-lingual content.

The key of the corpora used in this study is authentic language. The corpus approach (Biber, Conrad & Reppen, 1998:4) is used because it is empirical, analysing the actual patterns of language use. The corpora used in this study are composed of written text, a sample of Facebook public and personal posts, class lectures I give to my students as users of machine translation applications and a random sample of corpora and online texts. The main purpose is to determine how the various linguistic patterns and usage of corpora (a particular word, sound, text, syntactic, or morphological constructions) result in inaccurate machine translation.

Google (translate) and Facebook (see translation)

Computer technology has been applied in technical translation in order to improve speed and cost of translation (Trujillo, 1999). Translation with the aid of machines can be faster than manual translation and can reduce the cost of translation. In addition, the use of machine translation (MT) can result in improvements in quality, particularly in the use of consistent terminology within a scientific text or for a specific domain.
Google (translate) and Facebook (see translation) are the examples of the most currently used multilingual interne-based machine translation applications. The growing influence of the Internet has been reflected in the appearance MT applications like Google and Facebook investments in the area of Machine Translation to enhance their services to the Arab world users. In April 2006, Google (translate) was launched and originally created by Franz Josef Och with a statistical machine translation engine (Franz, 2006).

Google (translate) does not apply grammatical rules, since its algorithms are based on statistical analysis rather than traditional rule-based analysis. Google (translate) does not translate from one language into another (L1 to L2). Instead, it often translates first into English and then into the target language (L1, En, L2) (Christian et al, 2016). Some languages produce better results than others. Google (translate) supports 103 languages. English to and from Arabic was launched in April 28, 2006.

Facebook, a basically social networking service, was founded 13 years ago by Mark Zuckerberg. Its service covers the whole world except three blocking countries. Facebook’s latest update does not provide the service (see translation button). The following figures show how these applications work:

Figure 1: (Facebook “see translation” App)
It is apparent that translation plays a substantial role in human communication. There have been many translating strategies and types of equivalence. Some scholars prefer word-for-word translation, whereas others select sentence for sentence translation. A translator carries the burdens of conveying the meaning and effect of the text from one language to another.

To facilitate this task, several websites offering automatic / MT services, which translate not only corpus, sentences or even long document, were launched. Google Translation Service (GST) offers, state-of-the-art free translation service and works automatically without the intervention of human translators.

For several languages, one may see a speaker button near the translated text and by clicking this icon; one can hear a machine-generated spoken version of the translation. The Google translator allows translating whole documents, for example, in the form of PDF, TXT, DOC, PPT, XLS or RTF, or even images by just clicking the “translate a document” link and submitting a file without the need for copying and pasting large blocks of text (Sternby et.al., 2009).
A machine translation (MT) system is actually based on descriptions of both the source language corpus and the target language corpus at all levels: algorithm, formal grammars and vocabularies to produce translations. This machine translation process is based on the following steps (Eynde, 2015: 4):

- Analyze – source language text based on vocabulary, morphological and syntactical analysis
- Conversion (translation of source text to target text)
- Synthesis – creation of text for target language based on syntactical and morphological appearance of text.

All these steps in machine translation system may be interrelated and/or absent.

**Literature Review**

For systems developed during the 1990s the main sources are the survey of techniques by Trujillo (1999), the biennial “MT Summit” conferences and the numerous annual conferences and workshops for MT (e.g. Farwell et al. 1998, White 2000), computational linguistics (principally the Coling and ACL conferences), artificial intelligence, and information retrieval. In addition to other studies devoted to Arabic to English Machine Translation by Salem, Hensman and B. Nolan (2008).

The area of applying machine translation analysis to applications like Google translate and Facebook see translation is still a virgin territory.

MACHINE TRANSLATION (MT) of Arabic Corpora

With the late technological advancements in MT, Arabic has received attention in order to automate Arabic translations (Farghaly et al., 2009). The accuracy of human translation is very far from MT, as MTA does not guarantee that punctuation and spelling are one hundred percent precise. The corpus provided here is a sample which is frequently used and given to students as an example of simple sentences with complex and various layers of meaning and word orders. Google and Facebook are chosen because they provide freely available translations.

Therefore, there are issues and errors in MT, resulting in confusion regarding (word order) who is doing what or to whom or who reports to whom etc., see the following examples:

Arabic has a variety in word orders; this is a very urging problem in MTA due to the infinite prospects to express one sentence in Arabic.

1. V +N + N  2. N+V + N

This means that the subject and the object of each sentence have to be precisely identified. Table1 shows this challenge:

<table>
<thead>
<tr>
<th>SL (Arabic)</th>
<th>Gloss</th>
<th>MT</th>
<th>TL (English)</th>
</tr>
</thead>
<tbody>
<tr>
<td>يُحب محمد علي</td>
<td>Yu: Hib mohamed ‘ali</td>
<td>He loves Mohamed Ali</td>
<td>Mohamed Loves Ali</td>
</tr>
<tr>
<td>يُحب علي محمد</td>
<td>Yu: Hib ‘ali mohamed</td>
<td>I love Ali Mohamed</td>
<td>Mohamed Loves Ali</td>
</tr>
</tbody>
</table>
Table 1

The Difference between both examples is in the position of the actor. In example (1) the actor is the first argument of the verb.

In example (2) the actor is the second argument of the verb. Both sentences have the same meaning.

**Study Questions**

Through MT of a sample of Arabic and English variety of lexis, phrases and texts, the study tries to answer the following questions:

a. How efficient/deficient is MTA parallel corpora (target language output)?

b. What are the linguistic, grammatical, morphological, and contextual patterns associated with MTA inaccuracies?

c. What are the common errors of MTS?

d. How does MTS application functionally work?

**Sample of Corpus Analysis**

When translating Arabic, Machine Translation Applications have some semantic challenges, lexical choice mismatches or lexical and structural ambiguities. My corpora are presented according to prior set requirements, criteria with the purpose of showing the various linguistic patterns of Arabic corpora that pose a problem in MT. They include (lexis, phrases, concordances, and long sentences or utterances) See the following examples:

The screenshots below show how (Google translate application) or GTA translated the Arabic corpus ‘aqd into English. The Arabic (n)عِقدُ in English would be translated in this context as necklace.

In screenshot 1, the SL عِقدُ is typed without adding diacritics or the definite article آل and the result is inappropriate TL equivalent.

**Screenshot (1):**
In screenshot 1, the SL عقد is typed without adding any inflectional diacritics and the result was inappropriate TL equivalent.

**Screenshot (2):**

<table>
<thead>
<tr>
<th>Arabic – detected</th>
<th>اعطيتها عقدًا لكي تلبسه</th>
</tr>
</thead>
<tbody>
<tr>
<td>aetituha eaqdaan likay talbasah</td>
<td></td>
</tr>
<tr>
<td>I gave her a contract to wear</td>
<td></td>
</tr>
</tbody>
</table>

In screenshot 2, the SL عقدا is diacritised the result was inappropriate TL equivalent.

**Screenshot (3):**

<table>
<thead>
<tr>
<th>Arabic – detected</th>
<th>اعطيتها العقد كي تلبسه</th>
</tr>
</thead>
<tbody>
<tr>
<td>aetituha aleaqd kay talbisuh</td>
<td></td>
</tr>
<tr>
<td>Did you mean اعطيتها العقد كي تلبسه؟</td>
<td></td>
</tr>
<tr>
<td>I gave her a contract to wear</td>
<td></td>
</tr>
</tbody>
</table>

In screenshot 3, Inflectional diacritics “؟لafiaa لاام” (ال) of definiteness are added to the SL عقد and still we have the same misplaced TL equivalent.

**Corpus Source**

This paper purports to analyze the linguistic properties of a sample of SL corpora that result in inaccurate machine translation and improper TL parallel corpora. The Corpora are taken from some of the assignments I give to my university students during translation classes and various publically and personally shared Facebook posts. This papers shows how MT users, Facebook (see translation) and Google (translate) users apply machine translation to produce a parallel corpora. The sample of analysis does not provide imaginary idealized examples.

**Aims and Methodology**

This paper aims at tackling some of the linguistic problems facing the better utilization of Arabic Language on Internet based Machine Translation Applications (MTA). This paper presents a linguistic analysis of some of the problems resulting from the use of applications of MT,
namely Facebook (see translation) and Google (translate). The paper is applied to a selected sample of corpora comprised of formal and informal Arabic lexis, texts and Facebook posts. The sample is randomly selected to show the various linguistic patterns of Arabic and English languages. The corpora are presented in the form of screenshots taken from the MTAs webpages. All screenshots of the corpora collected are shown in the Appendix. The paper also discusses the challenges facing the Arabic language on internet based MTA; namely, Google (translate) and Facebook (see translation). In this paper, the impact of both non or partial diacritization on machine translation (MT).

The Challenges of Arabic Corpora to MT

An intact translation is one from which the TL parallel corpora readers can ably recognize the correct meaning of the SL corpus or text. Recent Machine Translation Application systems still make many blatant inaccuracies of meaning. Here are some of the issues of MTA to Arabic:

1. The problem of finding equivalent TL corpus for many language pairs especially when the SL is morphologically opulent.
2. A Singular Feminine adjective is mostly formed only by suffixing “taa? marbuta” to the masculine adjective. For example “mudarris” is the feminine of “mudarrisa” (teacher). Notice the following screenshots of GTA English rendering of the Arabic word مدرسة in both examples:

In screenshot (4), the source text is a single word (root+ feminine ending) with no diacritics.

In screenshot (5), the same word occurs in a clear context. The source text is a complete sentence (V+S+O). In both cases, the result is an inaccurate translation.

Screenshot (4)  
Screenshot (5)
3. Lexical diacritics differentiate between two lexemes. For example, the diacritization difference between the lexemes (بَاحِثْ) bะHIө ‘researcher’ and (بَاحَثَ) bAHaө ‘negotiated’ distinguishes between the meanings of the word (lexical ambiguation) rather than their inflections.

The following screenshots show how GTA produced an imprecise translation:

Besides, Having a non-diacritically marked Arabic text results in lexical and morphological ambiguity. Clear examples of this problem can be found on Facebook pages like “Insan” "إنسان" in which there are multiple mistranslated Arabic quotes:

تشتد و تشتد ثم يأتي الفرجة

taᶴttad wa taᶴttad umma ya?ti alfarraƷ
The word farraƷ is mistranslated as “vulva” instead of “relief” or “ease” which is a fatal mistake in translation and a sharp violation of the meaning of the sentence.

4. The shadda (gemination) diacritic adds further meanings to the lexeme حَضَرَ HaDar ‘attend’ versus حَضّرَ HaDDar ‘prepared’. Notice the following three screenshots (8, 9, and 10) that show these English rendering of the two verbs by GTA.
5. Arabic has no copula verb “to be“.
In Arabic there is no copula verb “to be” [Abn-Aqeal, 2007]. The verb to be’ is understood and then predicate subject. The subject and the predicate have to be in the nominative case, an example is shown in the table below:

<table>
<thead>
<tr>
<th>Arabic</th>
<th>Gloss</th>
<th>Google Translation</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>محمد طالِب</td>
<td>Mohamed Tælib</td>
<td>Mohammed Taleb</td>
<td>Mohammed (is) a student</td>
</tr>
<tr>
<td>محمد مُعلم</td>
<td>Mohamed Mu’alim</td>
<td>Mohamed Maalim</td>
<td>Mohamed (is) a teacher</td>
</tr>
<tr>
<td>محمد عَالِم</td>
<td>Mohamed ‘a’ælim</td>
<td>Mohamed world</td>
<td>Mohamed (is) a scientist</td>
</tr>
<tr>
<td>محمد طبيب</td>
<td>Mohamed Tᵊbi:b</td>
<td>Mohammed doctor</td>
<td>Mohamed is a doctor</td>
</tr>
</tbody>
</table>

Ahmed attended his papers
Ahmed attended his luggage
Ahmed attended the conference
6. The Arabic Language has various attachable clitics which include conjunction proclitics, e.g. و ‘and’, particle proclitics, e.g., ل ‘to/for’, the definite article the, and the class of pronominal enclitics, e.g., هم ‘their/their’.

7. Arabic has the specific case of dual, whereas other languages move from the singular to the plural form directly. In Arabic, we need only to add two letters to the singular form to express the dual form. An example is given in table (3):

Table 3

<table>
<thead>
<tr>
<th>Source Language</th>
<th>Gloss</th>
<th>Google Translation</th>
<th>Target Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>باب</td>
<td>Bæb</td>
<td>Door</td>
<td>Door</td>
</tr>
<tr>
<td>بابين</td>
<td>Bæbain</td>
<td>Babin</td>
<td>Two doors</td>
</tr>
<tr>
<td>بابان</td>
<td>Babæn</td>
<td>Babæn</td>
<td>Two doors</td>
</tr>
<tr>
<td>جرسين</td>
<td>ʒarasayn</td>
<td>jursæn</td>
<td>Two bells</td>
</tr>
<tr>
<td>جرسان</td>
<td>ʒarasæn</td>
<td>Jrsan</td>
<td>Two bells</td>
</tr>
<tr>
<td>امرأتين</td>
<td>i’mra?atain</td>
<td>Two women</td>
<td>Two women</td>
</tr>
<tr>
<td>قلم</td>
<td>qᵊlᵊm</td>
<td>Pen</td>
<td>Pen</td>
</tr>
<tr>
<td>قلمين</td>
<td>Qalamain</td>
<td>Two pencils</td>
<td>Two pens- pencils</td>
</tr>
<tr>
<td>بقرة</td>
<td>Baqara</td>
<td>A cow</td>
<td>A cow</td>
</tr>
<tr>
<td>بقرتين</td>
<td>Baqaratain</td>
<td>Two cows</td>
<td>Two cows</td>
</tr>
<tr>
<td>بقرتان</td>
<td>Baqarta:n</td>
<td>Bqrtan</td>
<td>Two cows</td>
</tr>
<tr>
<td>زهرة</td>
<td>Zahra</td>
<td>A flower</td>
<td>flower</td>
</tr>
<tr>
<td>زهرتان</td>
<td>Zahrata:n</td>
<td>Flowers</td>
<td>Two flowers</td>
</tr>
<tr>
<td>زهرتين</td>
<td>Zahratain</td>
<td>z-hrty</td>
<td>Two flowers</td>
</tr>
<tr>
<td>شجرتين</td>
<td>ʃajarain</td>
<td>Shjrtyn</td>
<td>Two trees</td>
</tr>
<tr>
<td>شجرتان</td>
<td>ʃajara:n</td>
<td>Wood</td>
<td>Two trees</td>
</tr>
</tbody>
</table>

9. The variety of the Arabic corpora used by Facebookers and Googlers does not include diacritics. However, diacritical markers are extremely
useful for clear readability and comprehension. Having a non-
diacritically marked Arabic corpora results in lexical and morphological
ambiguity.
10. Arabic diacritics are vowelization marks and usually absent (Schlippe,
et al 2008:271). “shadda” is the only diacritic which appears in several
modern Arabic scripts (Ibid). Native speakers distinguish the right
pronunciation and the correct meaning of a word without diacritic marks
by considering the context and the position of the word in a sentence.
11. There are four types of spelling errors:
1) word separation errors; 2) misspellings;
3) wrong capitalization; 4) wrong punctuation.

Analysis
The following table of analysis is based on a comparison of MT and
human Translation of a limited sample of random and frequently shared
corpora associated with many Facebook pages and electronically stored
corpus. Screenshots of the following corpora including the full electronic
text of each corpus and its MT “parallel corpus” are provided in the
appendix.
Table (4) shows the source language corpora and their machine
translation parallel corpora as provided by Google (translate) or
Facebook (see translation) in addition to a gloss for the corpora. A human
translation of the corpora is also provided in the table. The corpora chosen
here serve to answer the question of this paper:
What are the linguistic, grammatical, morphological, and contextual
patterns associated with MTA inaccuracies?
Answers of this question are shown in the analysis and notes column
(Table 4).
### Table 4

<table>
<thead>
<tr>
<th>SL Corpora</th>
<th>Gloss</th>
<th>MT Parallel Corpora</th>
<th>TL Parallel Corpora</th>
<th>Analysis &amp; Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>يروي قلبي</td>
<td>Yarwi: qalibi</td>
<td>Tells</td>
<td>(lit.) quench my thirst or warm my heart</td>
<td>MTA is unable to predict context sensitive word connotation. This is a concordance corpus whose meaning is not literal. It means warmed my heart. MTA lacks recognition of the concordance corpus meaning.</td>
</tr>
<tr>
<td>لتمحوها</td>
<td>litamHu:ha</td>
<td>To tamhouha</td>
<td>To be abolished, rubbed out, erased</td>
<td>MTA produces a Transliteration not translation. This is a morphologically complex corpus. MTA is unable to recognize the SL inflection. Multiple inflections added to the stem maHa:</td>
</tr>
<tr>
<td>لا تبتلينا</td>
<td>La tabta:li:na</td>
<td>Tptlina</td>
<td>Don’t let us be sorely tired by – afflicted with</td>
<td>MTA produced a Transliteration rather than translation. MTA is unable to recognize negative forms. MTA is unable to recognize multiple inflections.</td>
</tr>
<tr>
<td>هيحاسينا</td>
<td>haiHasibna</td>
<td>Hahasapena</td>
<td>Judge, consider</td>
<td>The corpora are examples of informal Arabic forms.</td>
</tr>
<tr>
<td>بتطلع</td>
<td>bitTala’</td>
<td>Aspira-tion</td>
<td>Bring out</td>
<td>MTA is unable to recognize Arabic varieties.</td>
</tr>
</tbody>
</table>

Misplacement of lexical
<table>
<thead>
<tr>
<th>Arabic</th>
<th>English</th>
<th>Equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>ممكن</td>
<td>mumkin</td>
<td>possible</td>
</tr>
<tr>
<td>تجيبلي</td>
<td>Tigi:bli</td>
<td>tjabla</td>
</tr>
<tr>
<td>مبتلبسوش</td>
<td>Mabtilbi:-su:ʃ</td>
<td>mbtalp-suc</td>
</tr>
<tr>
<td>أوجاعكم</td>
<td>?aw3aa’a-kum</td>
<td>auajje-km</td>
</tr>
<tr>
<td>بسجدة</td>
<td>bisa3da</td>
<td>bsgda</td>
</tr>
<tr>
<td>ستصيبك</td>
<td>satusi:b^k</td>
<td>stsepk</td>
</tr>
<tr>
<td>تحسب</td>
<td>taHsab</td>
<td>calculat-ed</td>
</tr>
<tr>
<td>ماشيه</td>
<td>maJa</td>
<td>cattle</td>
</tr>
<tr>
<td>بجد</td>
<td>bi3ad</td>
<td>hard</td>
</tr>
<tr>
<td>مش</td>
<td>meʩ</td>
<td>Mesh</td>
</tr>
<tr>
<td>عشان</td>
<td>‘a’an</td>
<td>Ashan</td>
</tr>
<tr>
<td>وتشيله</td>
<td>W t/i:luh</td>
<td>and chilh</td>
</tr>
<tr>
<td>تاني</td>
<td>ta:ni</td>
<td>tani</td>
</tr>
<tr>
<td>كي تري احلام</td>
<td>ila ‘alHulm</td>
<td>إلى الحلم</td>
</tr>
<tr>
<td>Fair</td>
<td>‘aadil</td>
<td>عادل</td>
</tr>
</tbody>
</table>

MTA is unable to disambiguate different meanings.

MTA cannot predict Vowelization.

MTA cannot recognize multiple inflections.

MTA produced a Transliteration not a translation.

This corpus is an informal negative form.

MTA is unable to recognize informal Arabic or language variations.

MTA lacks recognition or prediction of the SL corpus syntactic category and verb forms.
<table>
<thead>
<tr>
<th>Arabic</th>
<th>English</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ma’a ju: معكم</td>
<td>Ma’a ju: معك</td>
<td>MTA is unable to recognize semantic connotations of a corpus</td>
</tr>
<tr>
<td>Tadaa’ ?asfal تدع أسفل</td>
<td>Tadaa’ تخذه</td>
<td>MTA cannot predict the semantic reference of corpora concordances (phrasal verbs)</td>
</tr>
<tr>
<td>qaDa:a? قضاء</td>
<td>qaDa:a? أمضى - قضى</td>
<td>MTA is unable to recognize the function of lexicogrammar (relating the preposition to the verb)</td>
</tr>
<tr>
<td>fa’aarib Drinker</td>
<td>fa’aarib Fsharb</td>
<td>MTA is unable to recognize word’s syntactic category and function in a certain context</td>
</tr>
<tr>
<td>?aarib Drinker-drinking</td>
<td>?aarib Door</td>
<td>MTA is unable to recognize the function of lexicogrammar (relating the preposition to the verb)</td>
</tr>
<tr>
<td>Baaban Spokesman</td>
<td>Baaban Basem Morsi (Name) Spokesman</td>
<td>MTA is unable to recognize multiple semantic connotations of a corpus</td>
</tr>
<tr>
<td>Sharp Door</td>
<td>Sharp Door</td>
<td>MTA is unable to recognize the SL corpus inflections and complex morphology</td>
</tr>
<tr>
<td>Baba Spokesman</td>
<td>Baba Basem Morsi (Name of a person) Spokesman</td>
<td>Multiple inflections added to the stem ?ariba are not recognized by MTA</td>
</tr>
<tr>
<td>Morsi Spokesman</td>
<td>Morsi Spokesman</td>
<td>MTA is unable to recognize diacritics “tanweeen”</td>
</tr>
<tr>
<td>باسم مرسي</td>
<td>باسم مرسي</td>
<td>MTA cannot recognize context and syntactic category</td>
</tr>
</tbody>
</table>
Conclusion and Results
Through this limited sample of corpus of the MT of Arabic and English corpora provided in the study in form of screenshot or tables of analysis, it becomes very clear that there is a serious open problem related to MTs recognizing grammatical, morphological, syntactic and contextual patterns of the corpora (as noted and explained in table 4).

1. Machine Translation applications are not robust across different types of data, performing poorly on corpora whose underlying properties differ from those of the translation application data.
2. Most MT systems inaccuracies occur when dealing with morphologically rich Arabic corpora. The MT application will not generate word forms that they have not observed.
3. The corpora represent various linguistic patterns which generate inadequate and unacceptable MTA renderings. The corpora prove that the translation generated by GTA and FSTA is inadequate and unsatisfactory form linguistic point of view. These improper generated results give a negative impression of these online applications.
4. The transliteration forms provided by Google and Facebook are not accurate.
5. Undoubtedly the various flaws of GTA and FSTA shown in the table (4) prove that they are far from replacing professional translators.
6. Users of MT should use their common sense, which may resolve translation ambiguities.
7. These translation services may provide us with database of parallel corpora or root words.
8. Arabic dual word form is mostly unrecognized in GTA and FSTA.
9. These translation services should focus on providing multiple parallel corpora to polysemous word input that has multiple possibilities of translation.
10. These translation services should update their input methods database, and add diacritical markers and context sensitive recognition systems for the Arabic language.
11. Morphology reveals a number of translational problems, not all of which are the same type as shown in the table of analysis.
12. Machine translation that does not process language’s spirit can only turn put a corpse of parallel corpus.
References


Appendix
Screenshots of the Corpora Collected

1-

2-
3-

Google Translate

لله نعمة

4-

Lord, we do not hate it Tptlina defect in others

5-

هل تقصد
ربنا هبها سباع السعي مثا النتيجة #اسمي

6-

Forget Auajjaekm Bsgda long to God, prayer medicine pure hearts
7-

في ناس ماشي يببدأ نخاف من الله الناس يقول عليه بيبي أكثر من خوفنا من الله نينا يقول عليه حرام

fa nas mashiyh bimabda nakhaf min alla al...

8-

People in the cattle principle Elly afraid of people say it Aktar defect from our fear of our Lord Elly says it is haram

9-

Comfort Elly hard you fluidity every thing on our Lord
Fair is my love, when her fair golden hairs
With the loose wind ye waving chance to mark:
Fair, when the rose in her red cheeks appears,
Or in her eyes the fire of love does spark:

لا تدع أصدقاء أسفل عندما يحتاجون لك أكثر
la tade 'asdiqa’ 'asfal eindama yahtajun lak 'akthar
Cute relationship

When someone gets angry with you and says, “I will never talk to you” and later comes back to you to inform, „I AM STILL ANGRY.”
A Linguistic Analysis of the Problems of Facebook (see translation) and Google (translate)

God's gonna open baba I think
The severity of desperation, he didn't create with key..
#Polite
Rate this translation

Hug God every morning... and take him for a reasons... trust that everything good face
And we have to appreciate this good... and as we help ourselves and other.
#Goodmorning
#Polite
Rate this translation

I'm not LAZY... I'm just spending a little more time wondering what I should do first...

أنا لست كسول ... أنا مجرد فضفاض المزيد من الوقت يتساءل ما يجب أن أقوم به لأول مرة ...
‘ana last kusul ... ‘ana mjrd qada’ almaizid min alwaqt yatasa’al ma yjb ‘an ‘aqum bih ll’awwal marr ...
حببيه
باسم مرسي
صاحب وانتيمه طلع حلاق 😊😊

Ahieddiaier
Morsi spokesman 😊😊
Antimh owner and boarded the small barber 😊😊