

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

دراسة لغوية تحليلية لمشكلات الترجمة الآلية لتطبيقي فيسبوك وجوجل
لعينة من المتون

Dr. Heba Abdelraheim Ibrahim Alkady
Lecturer in Linguistics & Translation Studies
Faculty of Arts - South Valley University

د هبة عبدالرحيم إبراهيم القاضي
مدرس اللغويات والترجمة
كلية الاداب – جامعة جنوب الوادي

**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 لعينه من المتون

ملخص البحث

تعتبر الترجمة الآلية إحدى صور الابتكار في الصناعات اللغوية الحاسوبية وتمثل تقنية مذهلة في مجال دراسات وعلم الترجمة. يمثل هذا البحث محاولة التعرف أكثر على وسائل وتطبيقات مُستحدثه شائعة بين مستخدمي تطبيقات الترجمة العربية والانجليزية بمنظور آلي جديد وتحديد 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 الي 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 والدلالية للمفردات التي تمثل صعوبة لتطبيقات الترجمة الآلية اوعلامات التشكيل الا أنها مرجع ووسيلة سريعة يستخدمها كثير من الأفراد والمؤسسات في ترجمة المفردات والنصوص من وإلى الإنجليزية. حيث لوحظ ازدياد إقبال المستخدمين علي تطبيقات الترجمة الآلية مثل فيسبوك وجوجل. تهدف الدراسة إلي تقديم تحليل لغوي لبعض المشكلات الشائعة التي تواجه مستخدمي تطبيقي فيسبوك وجوجل تحديدا و لا 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 internet-based machine translation applications. The growing influence of the Internet has been reflected in the appearance of 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)

How do I get the Translate Facebook app?

Volunteer translators help make Facebook available in new languages, and can help improve the translations for existing languages.

Anyone who wants to bring their language to Facebook or improve the current translations on Facebook can be a translator. People who use Facebook in your language can see your approved translations.

Get the Translate Facebook App

To get the translate Facebook App:

- 1 From the main page of your app, tap **Search** at the top.
- 2 Type in "translate facebook" and tap **Search**.
- 3 Tap **Apps > Translate Facebook > Open**.
- 4 Select the language you want to translate into. We recommend using Facebook in the same language that you're translating

Figure 2 (Google “Translate” Homepage)

Google Translate homepage	
Type of site	Machine translation
Available in	103 languages, see below
Owner	Google
Website	translate.google.com 
Commercial	Yes
Registration	Optional
Users	Over 200+ million people daily
Launched	April 28, 2006; 10 years ago (as statistical machine translation) ^[1] November 15, 2016; 3 months ago (as neural machine translation) ^[2]
Current status	Active

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

Several studies have been conducted in the field of Arabic Language used online. Much research started with studies focusing on Arabic data input, character set analyzers, identification systems, search engines, machine translation and Arabic content online. Many of these efforts had contributed significantly to the field. Among these unique studies are the works of Beesley (1998) “Arabic Morphological Analysis on the Internet”, Sternby, Morwing, Andersson and Friberg (2009) “On-Line Arabic handwriting recognition with templates”, Selamat and Ng (2011) “Arabic script web page language identifications using decision tree neural networks,” Shen and Khalifa (2009) “Facebook usage among Arabic college students: preliminary findings on gender differences”, Zantout and Guessoum (2001) “An Automatic English-Arabic HTML page translation system”, Sanan, Rammal, and Zreik, “Internet Arabic search engines studies”. There have been plentiful efforts studying the effect of applying morphological processing (Hui, 1998; Habash and Sadat, 2006; Oflazer and Durgar El-Kahlout, 2007; Badr et al., 2008). Error analysis of MT using an open-source tool for error analysis of natural language processing tasks targeting morphologically rich languages was conducted by El Kholy and Habash, 2011. Habash and others have interesting works on Arabic Morphological Representations in Machine Translation (2004, 2005, 2006). The general history of MT is covered and updated by by Hutchins (1986, 1988, 1994, 1999, and 2001).

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:

	SL (Arabic)	Gloss	MT	TL (English)
1	يُحِبُّ مُحَمَّدٌ عَلِيَّ	Yu: Hib muhamed ‘ali	He loves Mohamed Ali	Mohamed Loves Ali
2	يُحِبُّ عَلِيٌّ مُحَمَّدًا	Yu: Hib ‘ali muhamed	I love Ali Mohamed	Mohamed Loves Ali

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:

- How efficient/deficient is MTA parallel corpora (target language output)?
- What are the linguistic, grammatical, morphological, and contextual patterns associated with MTA inaccuracies?
- What are the common errors of MTS?
- 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 (ن)عقد in English would be translated in this context as necklace.

In screenshot 1, the SL عقد is typed without adding diacritics or the definite article ال 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):



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

Screenshot (3):



In screenshot 3, Inflectional diacritics “ ?lif laam “ (ال) of definiteness are added to the SL عِقْدًا 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 “*mudarrisa*” is the feminine of “*mudarris*” (teacher). Notice the following screenshots of GTA English rendering of the Arabic word *مُدْرِسَة* in both examples:



Screenshot (4)

Screenshot (5)

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.

3. Lexical diacritics differentiate between two lexemes. For example, the diacritization difference between the lexemes (بَاحِثٌ) bəHIθ ‘researcher’ and (بَاخِثٌ) bAHaθa ‘negotiated’ distinguishes between the meanings of the word (lexical ambiguity) rather than their inflections.

The following screenshots show how GTA produced an imprecise translation:



Researcher flag

Screenshot (6)

Researcher matter with the government of President

Screenshot (7)

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 الفرج

taʔttad wa taʔttad ʊmma yaʔti 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 حَضَّرَ HaDDar ‘prepared’. Notice the following three screenshots (8, 9, and 10) that show these English rendering of the two verbs by GTA.

Arabic - detected

Arabic - detected

hadar 'ahmad 'awraqih حضر أحمد أوراقه

hadar 'ahmad aimtaeath حضر أحمد امتعته

Ahmed attended his papers Ahmed attended his luggage

Screenshot (8) Screenshot (9)

hadar 'ahmad almutamar حضر أحمد المؤتمر

Ahmed attended the conference

Screenshot (10)

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 2

Arabic	Gloss	Google Translation	English
محمد طاب	Mohamed Tælib	Mohammed Taleb	Mohammed (is) a student
محمد مُعَلِّم	Mohamed Mu'alim	Mohamed Maalim	Mohamed (is) a teacher
محمد عَالِم	Mohamed 'a'ælim	Mohamed world	Mohamed (is) a scientist
محمد طَيِّب	Mohamed T'bi:b	Mohammed doctor	Mohamed is a doctor

6. The Arabic Language has various attachable clitics which include conjunction proclitics, e.g. و w ‘and’, particle proclitics, e.g., ل l+ ‘to/for’, the definite article ال Al ‘the’, and the class of pronominal enclitics, e.g., هم hum ‘their/them’.

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

Source Language	Gloss	Google Translation	Target Language
باب	Bæb	Door	Door
بابين	Bæbain	Babin	Two doors
بابان	Babæ:n	Babæ:n	Two doors
جرسين	ʒarasayn	jursæ:n	Two bells
جرسان	ʒarasæ:n	Jrsan	Two bells
امرأتين	iʔmraʔatain	Two women	Two women
قلم	qʔlʔm	Pen	Pen
قلمين	Qalamain	Two pencils	Two pens- pencils
بقرة	Baqara	A cow	A cow
بقرتين	Baqaratain	Two cows	Two cows
بقرتان	Baqarta:n	Bqrtan	Two cows
زهرة	Zahra	A flower	flower
زهرتان	Zahrata:n	Flowers	Two flowers
زهرتين	Zahratain	z-hrty:n	Two flowers
شجرتين	ʃazarain	Shjrty:n	Two trees
شجرتان	ʃazara:n	Wood	Two trees

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

SL Corpora	Gloss	MT Parallel Corpora	TL Parallel Corpora	Analysis & Notes
يروى قلبي	Yarwi: qalibi	Tells	(lit.) quench my thirst or warm my heart	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
لتمحوها	litamHu:ha	To tamhouhha	To be abolished, rubbed out, erased	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:
لا تبتلينا	La tabta:li:na	Tptlina	Don't let us be sorely tired by – afflicted with	MTA produced a Transliteration rather than translation MTA is unable to recognize negative forms MTA is unable to recognize multiple inflections
هياسينا	haiHasibna	Hahasa- pena	Judge, consider	The corpora are examples of informal Arabic forms MTA is unable to recognize Arabic varieties
بتطلع	bitTala'	Aspira-tion	Bring out	Misplacement of lexical

ممکن	mumkin	possible	May	Equivalent
تجیلي	Tigi:bli	tjabla	bring me	MTA is unable to disambiguate different meanings
مبتلبسوش	Mabtilbi:-su:ʃ	mbtalp-suc	You don't wear	MTA cannot predict Vowelization MTA cannot recognize multiple inflections
أوجاعكم	?aw3aa'a-kum	auajjae-km	Pains suffering	MTA produced a Transliteration not a translation
بسجدة	bisa3da	bsgda	With a prostration	
ستصيبك	satusi:b^k	stsepk	Bless with	
تحسب	taHsab	calculat-ed	Consider-think	
ماشيه	ma ʃia	cattle	believe	
بجد	biʒad	hard	real	
مش	meʃ	Mesh	Not	This corpus is an informal negative form
عشان	'aʃan	Ashan	because, for and or then	MTA is unable to recognize informal Arabic or language variations
وتشيله	W tʃi:luh	and chilh	Keep store	
تاني	ta:ni	tani	Again	
تري احلام Fair	ila 'alHulm 'aadil	الي الحلم عادل	To dream جميل – ناعم	MTA lacks recognition or prediction of the SL corpus syntactic category and verb forms.

with u	Ma'a ju:	مع u	معك	<p>MTA is unable to recognize multiple semantic connotations of a corpus</p> <p>MTA is unable to recognize new corpus forms ((abbreviations and typed characters)</p>
Let down	Tadaa' ?asfal	تدع أسفل	تخذل	<p>MTA cannot predict the semantic reference of certain corpora concordances (phrasal verbs)</p> <p>MTA is unable to recognize the function of lexicogrammar (relating the preposition to the verb)</p>
Spending	qaDa:a?	قضاء	أمضي- اقضي	<p>MTA is unable to locate word's syntactic category and function in a certain context</p>
فشارب	fa'aarib	Fsharb	And the drinker	<p>MTA is unable to recognize the SL corpus inflections and complex morphology</p>
شارب	ʃaarib	Sharp	Drinker-drinking	<p>Multiple inflections added to the stem ʃariba are not recognized by MTA</p>
بابا	Baaban	Baba	Door	<p>MTA is unable to recognize diacritics "tanween"</p>
باسم مرسى	Basem Morsi (Name)	Morsi Spokesman	Basem Morsi (Name of a person)	<p>MTA cannot recognize context and syntactic category</p>

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 MTAs 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 from 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 out a corpse of parallel corpus.

References

- Abn-Aqeal. (2007). *Sharah Abn-Aqeal ala'lfat Abn-Malek*. Dar Al'alem Ilmlaien.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use* (Cambridge Approaches to Linguistics). Cambridge: Cambridge University Press.
- Block D. (2004). "Globalization, transnational communication and the Internet" *Social and Human Sciences*, 6, 22-37.
- El Kholy, Ahmed and Nizar Habash. (2010). "Techniques for Arabic Morphological Detokenization and Orthographic Denormalization". *In Proc. of LREC'10, Valletta, Malta*.
- El Kholy, Ahmed and Nizar Habash. (2012). "Rich Morphology Generation Using Statistical Machine Translation". *In Proc. of INLG'12, Utica, IL*.
- Eynde, Frank Van. (2015). *Linguistic Issues in Machine Translation*. New York: Bloomsbury.
- Farghaly, A. and J. Snellart. (2003, September 23). "Intuitive Coding of the Arabic Lexicon", Systran, MT, Summit IX Workshop, *Machine Translation for Semitic Languages: Issues and Approaches*, , New Orleans, Louisiana, U.S.A.
- Nizar Habash. (2004). "Large Scale Lexeme based Arabic Morphological Generation". *In Proceedings of Traitement Automatique des Langues Naturelles (TALN-04)*. Fez, Morocco.
- Nizar Habash, Owen Rambow, and George Kiraz. (2005). "Morphological Analysis and Generation for Arabic Dialects". *In Proceedings of the Workshop on Computational Approaches to Semitic Languages at 43rd Meeting of the Association for Computational Linguistics (ACL'05)*, Ann Arbor, Michigan.
- Nizar Habash and Fatiha Sadat. (2006). "Arabic Preprocessing Schemes for Statistical Machine Translation". *In Proceedings of the 7th Meeting of the North American Chapter of the Association for Computational Linguistics/Human Language Technologies Conference (HLT-NAACL06)*, Brooklyn, New York.
- , RANBO, Owen. (2007). Arabic diacritisation through full morphological tagging. [online] [cit. 30-11-2009]. Available at: <www.aclweb.org/anthology/N/N07/N07-2014.pdf>.
- Och, Franz Josef and Hermann Ney. (2003). A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1):19-52.
- Franz Josef Och. (2006, April). "Statistical Machine Translation live". *Google Research Blog*. Google. Retrieved December 1, 2016.
- Shen and M. Khalifa. (2009). "Facebook usage among Arabic college students: preliminary findings on gender differences", *Research online*, university of Wollongong.
- Sternby, J. Morwing, J. Andersson and C. Friberg (2009, December). "On-Line Arabic handwriting recognition with templates", *Pattern Recognition*, Vol. 42, No. 12, pp. 3279-3286.

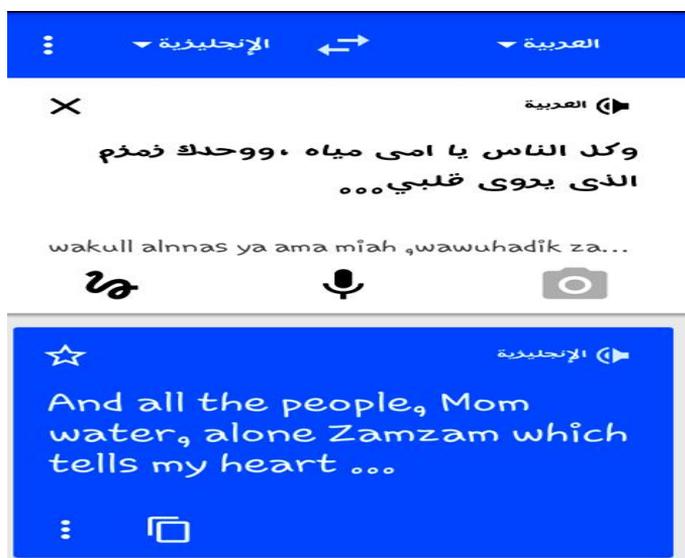
Thornborrow J. "Language and identity.(2004)." In, *Thomas, L. Wareing, S., Peccei, J.S. Thornborrow, J.(Eds) (2004), 158-172. Language, society and power: An introduction*, Routledge, London.

Trujillo, A. (1999). *Translation engines: techniques for machine translation*. London: Springer.

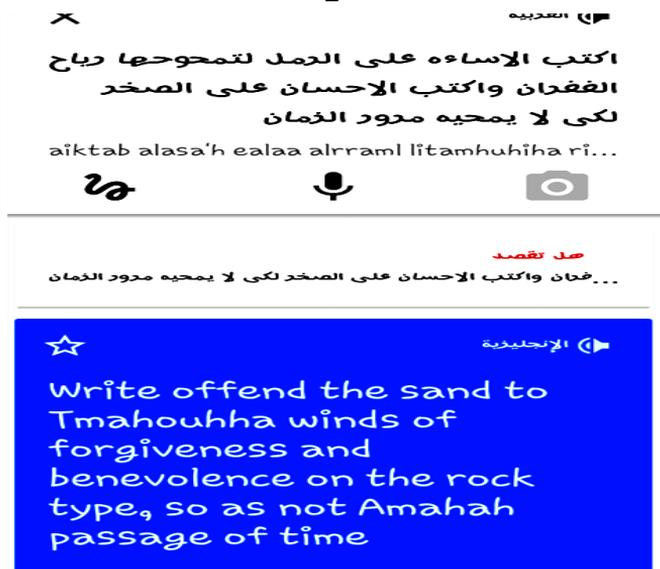
Appendix

Screenshots of the Corpora Collected

1-



2-



3-

Google Translate

الإنجليزية العربية

العربية العربية

يارب .. لا تبتلينا بعيب كرهناه في غيرنا

yarb .. la tabtalina beayb krehnah fi ghyrna

☆ الإنجليزية

Lord .. we do not hate it Tptlina defect in others

4-

هل تقصد

ربنا هيحاسبنا ع السعي مش النتيجة #اسعي

☆ الإنجليزية

Lord Hahaspena p quest mesh score
Knowledgeable

5-

العربية

انسوا اوجاعكم بسجده طويله لله
فالصلاه دواء نقي للقلوب

ansuu awjjaekum bisajdah tawiluh lillah f...

هد تقصد

انسوا اوجاعكم بسجده طويله لله بالصلاة دواء نقي للقلوب

☆ الإنجليزية

Forget Auajjaekm Bsgda
long to God, prayer
medicine pure hearts

6-

العربية الإنجليزية

العربية

ستصيبك رحمه الله من حيث لا تعلم
فى الأوقات التى تحسب فيها ان
النجاه مستحيله

satusibuk rahimah allah min hayth la tael...

Handwritten Arabic text: س

الإنجليزية

Stsepk God's mercy in terms
of you do not know at times
when calculated Surviving
an impossible

7-

فى ناس ماشيه بمبدأ نخاف من اللى
الناس تقول عليه عيب أكثر من خوفنا من
اللى ربنا يقول عليه حرام

fa nas mashyh bimabda nakhaf min allah al...

Handwritten Arabic text: س

صل تقصد
... بقول عليه عيب أكثر من خوفنا من اللى ربنا يقول عليه حرام

الإنجليزية

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

8-

العربية الإنجليزية

العربية

الراحة اللى بيدك انك تسبب كل حاجه على
ربنا

alrrahuh allah bijidd 'innak tusib kl hajh ealaa...

Handwritten Arabic text: س

الإنجليزية

Comfort Elly hard you
fluidity every thing on our
Lord

9-

8 مشاركات

مشاركة

English Poetry

١٤ ديسمبر، ٢٠١٢، الساعة ٥:٥٣ م

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:

العدل هو حبيبها, عندما شعر عادل الذهبي
مع " طليق الرياح تلوح فرصة للاحتفال:
عادل, عندما ارتفع في حدودها حمراء, يبدو
او في عينيها نار الحب شذرة:
تمت الترجمة من الإنجليزية

10-

Google Translate

العربية

↔

الإنجليزية

×

الإنجليزية

Don't let friends down when they
need you the most

🗨

🎤

📷

☆ العربية

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

☰

📄

11-



English Status



٦ نوفمبر، الساعة ١١:٣٨ ص ٠

Cute relationship

When someone gets angry with u and says,, "I will never talk to you"
and later comes back to u to inform....., "I AM STILL ANGRY..

علاقة جميلة

عندما شخص ما يغضب مع u و يقول,, "انا لن اتحدث معك"
وبعد ذلك يعود الى ان ابلغ..... " انا مازلت غاضبة..

تمت الترجمة من الإنجليزية

مشاركة واحدة



12-

فشارب #الخمير يصحو بعد
سكرته...
وشارب #العدس طول الشتاء
دفيان 🤔🤔🤔

#ömär

fsharib #alkhamar yashu baed sakratih.....



هل تقصد

...رب #العدس طوال الشتاء دفياني 🤔🤔🤔 #ömär



الإنجليزية

Fsharb # wine awakes
after Skrth ...
Sharp # lentils winter
along Dwian 🤔🤔🤔

Ömär

13-



مهديون

November 24 at 1:08am · 🌐

سيفتحُ الله باباً كنت تحسبه
من شدة اليأس لم يُخلق بمفتاحٍ ..
#مهديون

God's gonna open baba I think

The severity of desperation, he didn't create with key..

#Polite

⚙️ Rate this translation

14-



مهديون

November 24 at 8:30am · 🌐

استعن بالله كل صباح .. توكل عليه وخذ بالاسباب .. ثق ان كل ماتواجهه خير
وان علينا ان نقدر هذا الخير .. ونوظفه لمساعدته انفسنا والغير .
#صباح_الخير
#مهديون

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

15-

Google Translate

العربية ← الإنجليزية

× الإنجليزية ▶

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

☆ العربية ▶

أنا لست كسول ... أنا مجرد قضاء المزيد
من الوقت يتساءل ما يجب أن أقوم به
لأول مرة ...
'ana last kusul ... 'ana mjrd qada' almazid min
alwaqt yata'sal ma yjb 'an 'aqum bih li'awwal
marr ...

16-

احبيبييه
باسم مرسي 🙌🙌
صاحبه وانتيمه طلع حلاق 😊😊
😊😊😊



Ahieddiaier

Morsi spokesman 🙌🙌

Antimh owner and boarded the small barber 🐱🐱