Language Errors in Machine Translation of Encyclopedic Texts from English into Arabic: the case of Google Translate

Ahmad Muhammed Al-Samawi
College of Education, Sana'a University, Yemen
&
Al Ain University of Science & Technology, UAE

Abstract
Machine translation has facilitated the way for professional translators as well as ordinary people. Google Translate is undoubtedly the most popular machine translation program today. However, when translating texts from English into Arabic, the language errors that are produced by Google Translate are shocking, especially in encyclopedic texts. This article aimed at examining Google Translate errors in 10 encyclopedic texts of 10 sentences each. These texts were taken randomly from the free encyclopedia (Wikipedia) to represent 10 different academic disciplines. They were inputted into Google Translate to convert them from English into Arabic. The researcher traced **366** language errors in these texts (**3.66** errors per sentence). Based on error analysis, these errors were categorized into: syntactic errors (55 errors), grammatical errors (174 errors) and semantic errors (137 errors). These errors are signaling alarm to Google Translate team to improve the program of translation from English into Arabic and to find solutions to its current deficiency.

Keywords: English-Arabic translation, Google Translate errors, language errors, machine translation errors

Introduction

Technology has recently been playing a decisive role in daily activities; activities that are as small as momentary events or as big as annual plans. In other words, modern technology seems to intervene in virtually every human aspect including activities that thought to be solely performed by humans. Computers and internet, with no doubt, have influenced modern life and changed people's perception of life and literacy (Al-Samawi, 2012). Machine translation is one of the manifestations of such intervention of modern technology in human life.

Machine translation, as Hutchins and Somers (1992) explain, is the recent traditional and standard name for computerized systems responsible for the production of translations from one natural language into another, with or without human assistance. Similar to translation done by humans, machine translation does not simply involve substituting words in one language for another, but the application of complex linguistic knowledge: morphology (how words are built from smaller units of meaning), syntax (grammar), semantics (meaning), and understanding of concepts such as ambiguity (Diplo Foundation, 2011). Recently, machine translation has outrun the traditional machines, which were available to certain segments of people, and become publicly available online. Although machine translation has started in 1976 when Systran launched its first machine translation for the Commission of the European Communities (Selijan, Brkic, & Kucis, 2011), the first online free translation on the internet appeared in 1997 by Babel Fish using Systran technology (Aiken, Ghosh, Wee, & Vanjani, 2009a). According to Selijan et al. (2011, p. 331), "the use of online translation tools has increased in recent years, even among less widely spoken languages."

Arabic-English/ English-Arabic Machine Translation Programs

Google Translate: the most popular online tools for translation today is Google Translate, which was developed by Google and introduced in 2007 (Korosec, 2011). It is "a free translation service that provides instant translation between dozens of different languages" (Google Translate, 2013, par 1). Och (2006, par. 3) points out that Google launched an online version of its system for Arabic-English and English-Arabic. He states that "Arabic is a very challenging language to translate to and from: it requires a long-distance reordering of words and has a very rich morphology." Lately, Google Translate has been massively used by a wide spectrum of people: academics, students, novice translators, professional translators and so on.

How does Google Translate work? The best answer to such a question is the one that is posted by the Google Translate Team in Google Translate (2013, par. 2) website states:

> When Google Translate generates a translation, it looks for patterns in hundreds of millions of documents to help decide on the best translation for you. By detecting patterns in documents that have already been translated by human translators, Google Translate can make intelligent guesses as to what an appropriate translation should be. This process of seeking patterns in large amounts of text is called *statistical machine translation.*"

In the beginning, Google Translate used Systran, but in October, 2007, Google switched the translation system from Systran to its own machine translation system for all 25 language pairs available on the site though it has used its own system earlier in Arabic, Chinese, and Russian (Chitu, 2007; Schwartz, 2007; Korosec, 2011).

To improve the quality of translation and to facilitate quick translation, Google Translator Toolkit was launched in July 2009, which "is basically a collaborative web-based translation memory (TM) platform into which translators upload texts and submit them for translation" (par 5). In February 2013, Google Translate team announced the integration of Google's new input tools in Google Translate that expands the set of available input methods for many languages including Arabic (Chin, 2013a). In March 2013, Google Translate launched offline packages for Google Translate on Android with support for fifty languages, including Arabic (Jiang, 2013a) and in May 2013, Google Translate developed the phrase-book where the travelling users can get access to their favorite translated phrases. As Jiang (2013b) explains, Google Translate lets users save translations of phrases in a program called Phrasebook (Jiang, 2013b). Also in May, 2013, Kelman (2013, par. 4), one of Google Translate managers, reports that Google Translate has reached "70+ language milestone" while the quality isn't perfect. Moreover, as Chin (2013b) points, Google Translate improved the service of the paid YouTube video caption translation.

Arabic is one of 71 languages which Google Translate currently supports (Google Translate, 2013). At the word level, it provides alternatives in case a translation does not seem right. Such a technique is also used as a feedback which helps Google Translate improve the quality of machine translation. Besides, Google Translate suggests using Translator Toolkit (Google Translate, 2013). However, an overview of other programs developed for Arabic-English and English-Arabic machine translation seems to be necessary before dealing with the research problem.

ATA: the first Arabic software for machine translation was developed in the mid-90s of the last century, when ATA, a London-based software house specializing in Arabic business software, released Al-Mutarjim Al-Arabey (the first English-Arabic machine translation software on PCs and Macintosh computers). Such a dictionary was modernized in 2002 when ATA released Al Mutarjim™ Al Arabey v3.00 followed by Mutarjim*Net*™ v1.00 in 2003. In 2004, a new translation program called 'Arabic Memory Translation system XPro7' was launched, followed by a Beta version of the ATA Arabic search engine ALHOODHOOD in 2005. In January 2006, ATA launched a name Translation System, where names from different countries written in English are translated correctly into Arabic. Al-Wafi is another product of ATA Software. Many versions have been developed since the release of Al-Wafi Quick Dictionary v1.00 in February 2002 followed by Al-Wafi v3.00 and finally the Golden Al-Wafi v1.00 of which the new version was launched in 2007.

According to ATA Software Technology Limited (2013), ATA Software previewed in 1997 and a pre-release version of a revolutionary new piece of Arabic Text-To-Speech software, called Al-Natiq. Al-Natiq was presented at the Gitex '97 exhibition in Dubai, UAE. In 2000, ATA launched almisbar, an online translation service which became a valuable tool for hundreds of thousands of Arabic users. It was designed to provide free and instant translation of English websites and texts, and a handy bilingual dictionary. Lately, ATA Software has successfully installed a full local version of *almisbar*, an online English-into-Arabic translator, at the *Princess Nora bint Abdulrahman University* (PNU), Saudi Arabia. The translation system runs on a local server which is accessed in Saudi Arabia only by students using their PNU email addresses. The entire system is installed on the PNU servers completely independent of the ATA Software London servers.

Atlas: Atlas is one of the earliest machine translation programs from and into Arabic. It was developed by the FTC (First Trading Company) in Hong Kong (Atlas, 2013). The company has developed many versions of the program in the form of electronic, online and paper

dictionaries. Atlas electronic pocket dictionaries are more famous in the Arab world than other electronic dictionaries that deal with translation from and into Arabic. Many versions of ATLAS translators have been introduced such as the SM series and the SD series. The latest version of Atlas is the *Atlas Modern Dictionary*, which is made available online and *Atlas Dictionary L519*.

Babylon: Babylon has also contributed to machine translation from and into Arabic. According to Dictionaries and Encyclopedias (2013), Babylon Ltd. has developed 36 English-based proprietary dictionaries in 21 languages, including Arabic, and made them free of charge to users of the software. These dictionaries comprise between 60,000 to 200,000 terms, phrases, acronyms and abbreviations and are enabled with a morphological engine which facilitates the recognition of all inflected forms of single words and phrases, provides all forms of terms that include prefixes and extensions and supplies a solution for all formats of writing.

Research Problem

Although we live in the age of advanced technology, and machine translation has improved a lot since it was developed, the accuracy of machine translation output still witnesses great deficiency, especially when two languages of different linguistic systems are involved, as in the case of Arabic and English. By linguistic systems it is meant morphology, grammar, syntax, and semantics, orthography as well as style. The deficiency of machine translation can be observed in the results of inputting a corpus of texts from different disciplines into an automated translation program from one natural language to another. No one can deny that machine translation is superior in time and money saving to human translation. When the accuracy of the translation is compared to professional human translation, however, there is no doubt that the accuracy in the latter is much higher than that in the former. As Berner (2003, p. 10) states, "The accuracy of MT is much lower than competent human translation, but can be improved in certain ways."

Google Translate is one of the highest programs of machine translation in language pairs. Aiken, Park, Simmons & Lindblom (2009b) indicate that Google Translate works with 1640 language pairs including Arabic. Nevertheless, the accuracy of Google Translate is affected by the type of pairs in which translation takes place. "This is why translation accuracy will sometimes vary across languages" (Google Translate, 2013, par 2). As Google Translate Team puts it, "since the translations are generated by machines, not all translations will be perfect" (Google Translate, 2013, par 2). Or as Aiken, Park, Simmons & Lindblom (2009, p. 5) state, "Translation comprehension is still far from perfect because Google Translate's accuracy varies with sentence and vocabulary complexity and by language."

Arabic is a language that has a different linguistic system from English, consequently, machine translation between the two languages may not be as accurate as it may be between English and other Indo-European languages. Google Translate, and to less degree other machine translation programs, has tried to bridge the gap between the language systems through continuous improvement of the translation programs used as shown before. At the word and phrase level, translation programs from and into Arabic seem to do an excellent job, though a minor violation of the Arabic phrase structure is found in the output of some of these programs. A funny translation of the English phrase "CAUTION: WET FLOOR" using earlier version of Google Translate rendered: الحذر الكلمة الرطب '*al hathar al kalemah arratb*' (English: The word, the caution, the soft), while the appropriate equivalent phrase in Arabic is: تنبيه: ارضية مبلولة '*tunbeeh: ardheyah mubloolah*'. In the new version, however, the output depends on the case of the English letters. For example, if the whole phrase is written in uppercase letters, the result is: تنبيه: ارضية مبللة '*tunbeeh: ardheyah muballalah*'; if the letters are in lowercase, the result is: الحذر الطابق الرطب

'*al hathar at-taabaq arratb*'. Moreover, Al-Samawi (2013) has traced errors in promotional circulars and found that some of these errors were due to Google Translate performance of products names. At the sentence and paragraph level, however, these programs are still in need for improvement. This is not to mention the funny outcome of the texts translated. Such improvement should be based on real data taken from different disciplines in which deficiency of the translation is clear.

Error analysis of text translation is one of the techniques that reveal the weaknesses of machine translation programs and help programs developers find out solutions to their current problems. Unfortunately, most of the previous studies that tried to use error analysis in machine translation research were at the level of the single word or phrase. Like a rare bird, research on errors of machine translation at the text level may not be easy to find, especially in Arabic-English. Thus, the present article intends to examine the language errors produced by Google Translate when translating encyclopedic texts from English into Arabic.

Research objectives
The primary aim of the present study is to find the errors produced by Google Translate when translating encyclopedic texts of different academic disciplines from English into Arabic. The research will address the following objectives:
1. To identify the language errors produced by Google Translate in encyclopedic texts of different disciplines translated from English into Arabic.
2. To analyze such errors linguistically and classify them according to their types.
3. To evaluate their effect in corrupting the meanings at the phrase and sentence levels.

Research significance
The present article is a pioneer in its topic, (i.e., errors of machine translation of encyclopedic texts from English into Arabic). Previous studies have tackled these errors, but at the level of single words, phrases and simple sentences. The results of this article will be of significant value as they provide Google Translate with an authentic basis for improving its current machine translation program from English into Arabic. Moreover, the present study could be used significantly in future research that aims at contrasting linguistic systems of Arabic and English, which in turn, helps contrastive linguistic studies in general and Arabic-English contrastive studies in particular. Besides, it adds to the current literature in machine translation, which helps future research form certain theoretical considerations about Google Translate in particular and machine translation in general.

Research limitations
The present study is limited to Google Translate errors found in encyclopedic texts translated from English into Arabic. It does not tackle, however, the errors of other machine translation programs. Besides, it is limited to language errors at the text level in different encyclopedic texts of different disciplines. Evaluation of machine translation programs is also beyond the scope of the present study. It is also beyond the scope of the present study to compare the outputs of the different available machine translation programs with Google Translate.

Previous Studies

On machine translation

Historically, research in machine translation started immediately after World War 2 (Aiken and Vanjani, 2009). Although it has been a central issue in modern technology, the amount of research done about errors of machine translation does not match such an eminent status. Most of the research conducted in this regards was intended to compare programs of different developers or to measure the effectiveness of these programs. Recently, machine translation has received substantial research attention. But as Dhore, Dixit, & Sonwalker (2012) explained, the source language, in most cases, has been English and the target language is an Asian language. Machine translation evaluation has been carried using different techniques ranging from traditional methods such as BLEU (Bilingual Evaluation Understudy) and TER (Translation Error Rate) to quality perception by native human evaluations (Farrús, Ruiz Costa-Jussà, Mariño, & Rodríguez, 2010) and automatic evaluation metrics (Brkic, Mikulic, & Matetic, 2012).

In the study that was carried by Aiken et al. (2009a, par 4), they compared six web-based machine translation services and introduced "a new, locally developed multilingual electronic meeting system that provides automatic translation among 41 languages." They tested three concepts in text translation: comprehension, acceptability and meaning. As for meaning (the concept related to the present study), they indicate five results of text translation, ranging from conveying the exact original meaning to the deficiency in conveying the meaning at all.

Previous studies in machine translation differentiated between simple sentences and complex sentences. For example, Zervaki (2002) points out that in the case of simple sentences and SVO order, machine translation can produce acceptable terminology and syntax. However, in more complex sentences translations become incomprehensible.

On Google Translate

As mentioned earlier, Aiken et al. (2009b, p. 5) listed Google Translate as the highest machine in language pairs with 1640 language pairs including Arabic. However, they state that "Translation comprehension is still far from perfect because Google Translate*'s* accuracy varies with sentence and vocabulary complexity and by language." Resnik, Buzek, Hu, Kronrod, Quinn, & Bederson (2010, p. 136) examined the effect of targeted paraphrasing on improving machine translation of sentences from Chinese to English and found using targeted paraphrasing can significantly improve translation. They choose Google Translate for "its wide availability and the fact that it represents a state of the art baseline to beat." Selijan et al. (2011, p. 343) also remarked favorably on Google Translate saying that it "seems to be well trained and suitable for the translation of frequent expressions." Although they were optimistic that the use of a background terminology database of multiword expressions and/or translation memory database would probably improve results, especially translations of specific terms and idiomatic expressions, they complained that Google Translate does not perform well where language information is needed, such as gender agreement.

The study that was carried by ElShiekh (2012) is, perhaps, the closest in topic to the present study among research conducted on Google Translate. He ran an investigation exploring the nature of the translation process provided by Google Translate Service from English into Arabic and vice versa. However, his study used three types of texts: advertisements, literary styles and religious texts, and the translations were performed from Arabic into English and vice versa. The present study uses ten texts of different encyclopedic disciplines with one-direction

translation (i.e. from English into Arabic). Nevertheless, ElSheikh (ibid, p. 56) warned against taking the final translation of Google Translate by the average Internet users who are not professional translators "for granted even where it could be totally catastrophic." Similarly, Jamilah (2012) used the errors analysis of homonymous and polysemous word structure in Google Translate from Indonesian into English. She found that Google Translate program was not able yet to translate a hundred of homonymous and polysemous words that are created by the changes in the word function. Because Google Translate can create lexical ambiguity, she suggested to the language users not to rely only on Google Translate in translating any SL.

Farrús et al (2010, p. 172) compared Google Translate performance to the N-II (a machine translation program developed at the Technical University of Catalonia). They used BLEU and TER to evaluate the linguistic errors and compared them to a new human evaluation based on the expert knowledge method about the errors encountered at several linguistic levels: orthographic, morphological, lexical, semantic, and syntactic. They came to the conclusion that the lexical and the semantic levels "have more influence on the way how the human evaluators perceive the errors. In the same way, both lexical and semantic errors seem to be also consistent with the automatic evaluation measures BLEU and TER."

Though perhaps not directly related to the present study, a recent study was carried out by Balk, Chung, Chen, Trikalinos, & Kong (2013) where they assessed the accuracy of Google Translate to allow data extraction from10 articles published in 5 languages and the time needed by Google Translate for translating into English. They found that the length of time needed ranged between 5 minutes to one hour for almost all the articles, but the errors of translation were not examined.

On Arabic-English and English-Arabic machine translation

Aiken & Vanjani (2009) tested locally developed internet-based electronic group support system that automatically translates between 34 languages including Arabic. They indicate that although Arabic is not similar to English, the results provide good translation. As mentioned in the previous section, Aiken et al. (2009b) tested three concepts in machine translation of texts: comprehension, acceptability and meaning. They ranked Arabic in 38 among the forty languages used in both comprehension and acceptability, and 37 in meaning. Oweis (in Hujair, 2012) summarized the difficulties which Google Translate faces in translating texts from English into Arabic and vice versa as: (1) the direction of text from right to left, (2) masculine and feminine and Arabic sentence structure, and (3) singular and plural and numerical rules. To improve the quality of translation from English into Arabic and vice versa, Google has set up a workshop with Arab media and journalists to adopt about 6000 Arabic terms that correspond to technological terms (Oweis quoted in Hujair, 2012).

Al- Kabi et al (2013) compared the effectiveness of Google Translate and Babylon in translating well known sayings and English sentences into Arabic using BLEU. They found that Google Translate was better than Babylon in terms of precision of translation.

Research Methodology

The present research is a descriptive analytical study, using a combination of content analysis and error analysis techniques. According to Stemler (2001, par 1) "content analysis has been defined as a systematic, replicable technique for compressing many words of a text into fewer content categories based on explicit rules of coding." Error analysis is a technique that was developed in language studies during the second half of the twentieth century as a result of the

growing concept of interlanguage. Crystal (2008, p. 173) defines error analysis as a "technique for identifying, classifying, and systematically interpreting the unacceptable forms produced by someone learning a foreign language, using any of the principles and procedures provided by linguistics." Accordingly, ten randomly selected texts were used from which data was obtained. Error analysis was used as a primary technique for analyzing the language used, identifying and classifying errors in every sentence of the selected texts. As can be noticed, error analysis was mainly used for analyzing SL/FL learners' errors. In the present study, Google Translate is treated as a learner of Arabic as a Second/Foreign Language (ASL/AFL) with the reserved differences between human and machine. According to Larsen Freeman and Long (1991), error is a systematic deviation from the standard language.

Procedures

Disciplines and text
Ten English encyclopedic texts were selected randomly (main research sample) from the Free Encyclopedia (Wikipedia). The ten texts represent ten particular areas of knowledge (research population) taken from ten academic disciplines following Melville Dewey's decimal classification system, namely: philosophy, religion, media, education, linguistics, physics, technology, literature, geography, and history. Only the first ten sentences from each text were used as the actual sample of the research. The texts were of different topics as shown in Table (1).

Treatment. Each text then was inputted into Google Translate (2010); with the order to translate it into Arabic. The resulted translation was cut from Google Translate box and pasted on a separate sheet for each text. Because of the difference between Arabic and English systems of punctuation and writing, the English sentence was used as a unit of analysis (Appendix A). The revealed Arabic words in each sentence were calculated and added to form the total number of Arabic words in each text. The number of average words in every text was also calculated. The ten texts (100 sentences) contained 1,795 Arabic words. Texts were then arranged in ascending order (from the lowest to the highest).

Table 1: Texts according to the total number of words and their average

Text	Topic	Total number of words	Average words	Number of words in the longest sentence	Number of words in the shortest sentence
Philosophical Text	Epistemology	144	14.4	20	9
Religious Text	Hadith in Islam	226	22.6	39	8
Media Text	Journalism	244	24.4	40	13
Educational Text	Educational Assessment	150	15.0	22	19
Linguistic Text	Language and	201	20.1	28	13

	Culture				
Physical Text	Applied Physics	157	15.7	33	6
Technological Text	Software Programming	164	16.4	26	8
Literary Text	American Literature in 20th century	165	16.5	27	9
Geographic Text	Lithosphere	153	15.3	27	8
Historical Text	American History	191	19.1	32	10
Total Words		1,795	17.95	29.4	10.7

As can be seen from the table above, the media text contained the highest number of words (244 words) followed by the religious text (226 words). On the other hand, the philosophical text contained the lowest number of words (144) followed by the educational text (150). The longest sentence was found in the media text (40 words) followed by the religious text (39 words). In contrast, the shortest sentence was found in the physical text (6 words) followed by the religious, technological, and geographic texts (8 words in each).

Content analysis and error analysis
As mentioned earlier, content analysis is used in the present study along with error analysis. The error analysis technique was used according to the steps suggested by one of the most influential figures in error analysis (i.e. Stephen Pitt Corder). He lists five procedures for error analysis: Select a corpus of language, identify the errors in the corpus, classify such errors, explain them, and evaluate how serious they are (Corder, 1967). For the present study, the first four procedures are followed consecutively; the last is implied in the discussion of the results. However, the procedure went through two main phases.

The first phase involved the research scrutinizing through the texts; sentence by sentence and word by word to establish the categories of errors. 17 types of errors were revealed. Such types were then categorized into three main categories: syntactic errors, grammatical errors, and semantic errors. There were two types of syntactic errors and were given numbers 1 and 2. Grammatical errors, on the other hand, were sorted into seven types which were given numbers from 3 to 9. The last category that had more types was the semantic errors, in which 8 errors were traced in the translated texts. They were given numbers from 10 to 17.

For the sake of reliable judgment and categorization of errors, the Arabic texts and the categories were given to two PhD holders in Arabic language teaching at the rank of associate professors. They both agreed with the categories and made minor modifications to the list. Later, the categories and the types of errors were translated into English for the purpose of reporting the final research in English language (Table 2). A table in Arabic was then developed to be used as a worksheet in which categories of errors were written in Arabic. Again, for the purpose of reporting, the table was converted into English (Appendix B).

Table 2: Categories and types of error found in the translated texts

Error Category	No	Type of Error
Syntactic Errors	1	Starting with a nominal sentence in the place of a verbal sentence.
	2	Violating the whole phrase structure (Putting adjective before noun, Putting modifiers before modified terms
Grammar Errors	3	Using wrong form of the word (plural, the five verbs, five nouns, nouns and verbs inflections)
	4	Violating subject-verb agreement (masculine and feminine; singular, dual, and plural; first, second, and third person)
	5	Using a noun in place of a verb
	6	Using a verb in place of a noun
	7	Using wrong prepositions, articles, and particles
	8	Using definite article before genitives
	9	Omitting functional morphemes (i.e. prepositions, articles, conjunctions, pronouns, auxiliary verbs, deixis, etc.)
Semantic Errors	10	Using a wrong meaning of English homonyms
	11	Using words of ambiguous meaning
	12	Using terms that convey very different meaning
	13	Using unfamiliar words in place of collocations
	14	Using wrong reference and relative pronouns.
	15	Adding an unnecessary word, preposition, or article before a word
	16	Omitting necessary words or phrases
	17	Corrupting the meaning of the whole sentence

The second phase involved the researcher going over each text; sentence by sentence and word by word, labeling each error with numbers according to the list that was originated in the first stage. The total number of errors in each category and the total number of errors in general were then tallied up and recorded in the Arabic table. The results were then transferred into the English version (Appendix B).

Results

The total number of errors traced in all the translated texts was 366 errors. When comparing such a number to the number of the total words in all the texts examined, it represents about 20.3% of these words.

Table 3: Percentages of Errors to Total Words in Texts

Text	Total number of words in each text	Total number of errors in each text	Percentage of errors in each text to total errors in all texts	Percentage of errors to the total number of words in each text
Media Text	244	37	10.1%	15.2%
Religious Text	226	47	12.9%	20.8%
Philosophical Text	144	26	7.1%	18.1%
Linguistic Text	201	38	10.4%	18.9%
Physical Text	157	30	8.2%	19.1%
Literary Text	165	34	9.3%	20.6%
Geographic Text	153	35	9.6%	22.9%
Historical Text	191	46	12.6%	24.1%
Educational Text	150	37	10.1%	24.7%
Technological Text	164	36	9.8%	22%
Total	1,795	366	100%	-

Looking at Table (3) above, it can be noticed that the total number of errors was 366 compared to the total number of sentences (100), which means 3.6 errors per sentence. Meanwhile, the number of total errors compared to the number of words in each text revealed that the educational text crowned the list with 24.7%, followed by the historical text (24.1%), while the media text had the lowest percentage (15.2%). These percentages are indicators of the level of difficulty which Google Translate faces when translating different texts from English into Arabic.

On the other hand, the religious text had the highest frequencies of errors (47 each) that represent 12.9% of the total errors. The historical text came in the second place with 46 errors that represents 12.6%. The lowest number of errors was found in the philosophical text (26 errors) which represents 7.1% of all errors; then the physical text (30 errors; 8.2%).

Categorical error analysis revealed that most of the errors were grammatical errors (174 errors) forming 47.5% of the total errors (366) (Table 4). Meanwhile, semantic errors (Table 5) came in the second place (137 errors) forming 37.4% of the total errors. On the other hand, syntactic errors were the lowest in number (55 errors) representing 15.0% of the total errors (Table 6).

Within errors, error type 9 (*Omitting functional morphemes 'prepositions, articles, etc.*) had the highest frequency of occurrence (54 times) among all errors, which represents 14.8% of the total number of errors. Such a category had its mode in the educational text as the highest among all texts (13 times). Error type 15 (*Adding an unnecessary word, preposition, or article before a word*) came in the second place with 51 frequencies (13.9% of the total errors) and peaked in the media text as the highest (10 times); followed by error type 2 (*Violating the whole phrase structure*) with 50 frequencies (13.7% of the total errors) with the linguistic text at the top

of the list (9 times). The lowest type of errors was error type 6 (*Using a verb in place of a noun*) with 1 frequency, followed by error types 13 and 17 with 3 frequencies each (See Appendix B for a summary of all results).

At the category level, item no 2 (*Violating the whole phrase structure*) was the highest among syntactic errors. It counts 94.3% of the syntactic errors.

Table 4: Syntactic Errors

No	Types of Syntactic Errors	Field of text and frequency of errors										Total
		Media	Religious	Philosophical	Linguistic	Physical	Literary	Geographic	Historical	Educational	Technological	
1	Starting with a nominal sentence in the place of a verbal sentence.	1	1	0	1	0	0	0	1	0	1	5
2	Violating the whole phrase structure	4	3	5	9	5	7	5	6	3	3	50
Total		5	4	5	10	5	7	5	7	3	4	55

On the other hand, error type 9 in grammatical errors (o*mitting functional morphemes*) was the highest, representing 30.5% of the total errors in that category.

Table 5: Grammatical Errors

No	Types of Grammatical Errors	Field of text and frequency of errors										Total
		Media	Religious	Philosophical	Linguistic	Physical	Literary	Geographic	Historical	Educational	Technological	
3	Using wrong form of the word (plural, the five verbs, five nouns, nouns and verbs inflections)	7	2	1	2	3	1	7	4	3	11	41
4	Violating subject-verb agreement (masculine and feminine; singular, dual and plural; first, second, and third person)	0	2	1	3	2	4	4	4	1	1	22
5	Using a noun in place of a verb	1	1	0	2	3	3	2	1	0	0	13
6	Using a verb in place of a noun	0	0	0	1	0	0	0	0	0	0	1
7	Using wrong prepositions, articles, and particles	4	6	3	5	3	1	3	6	1	7	39
8	Using definite article before genitives	0	3	0	0	1	0	0	0	0	0	4
9	Omitting functional morphemes (prepositions, articles, etc)	4	11	5	3	2	5	2	5	13	4	54
Total		16	25	10	16	14	14	18	20	18	23	174

Error type 15 in semantic errors (a*dding an unnecessary word, preposition, or article before a word*) counts for 37.2% of the total semantic errors.

Table 6: Semantic errors

No	Types of Semantic Errors	Field of text and frequency of errors										Total
		Media	Religious	Philosophical	Linguistic	Physical	Literary	Geographic	Historical	Educational	Technological	
10	Using a wrong meaning of English homonyms	1	8	6	2	3	1	3	4	3	6	37
11	Using words of ambiguous meaning	0	0	3	0	0	3	0	0	0	0	6
12	Using terms that convey very different meaning	1	1	0	1	2	1	1	3	0	1	11
13	Using unfamiliar words in place of collocations	0	0	0	0	0	1	0	0	1	1	3
14	Using wrong reference and relative pronouns.	2	2	0	1	2	2	0	1	4	0	14
15	Adding an unnecessary word, preposition, or article before a word	10	3	1	5	4	4	8	9	6	1	51
16	Omitting necessary words or phrases	1	3	0	3	0	1	0	2	2	0	12
17	Corrupting the meaning of the whole sentence	1	1	1	0	0	0	0	0	0	0	3
Total		16	18	11	12	11	13	12	19	16	9	137

Discussion

The main objective of the present research was to discover the errors produced by Google Translate in translating encyclopedic texts of different academic disciplines from English into Arabic. Such an objective was achieved through four stages: selecting texts translated by Google translate from ten disciplines, identifying errors in these texts, classifying such errors, and reporting them. In general, 366 errors were traced in the 100 sentences used for analysis (average 3.66 errors in each sentence; 20.3 % of the total words) in the ten texts. Such a number depicts a considerable weight of errors in Google translation of texts from English into Arabic (Table 3). The results, in general, lend support to ElSheikh (2012) findings. However, they differ from ElShiekh's in that they are related to encyclopedic texts, rather than general texts and the

direction of translation is from English into Arabic while his study was from Arabic into English and vice versa. Moreover, the present study categorized and subcategorized errors in a way that makes an easy reference to these errors. Besides, he used short translation assignments given to students, while the texts of the present studies were taken from the Wikipedia (2013).

Syntactic errors

Starting with syntactic errors, the category *"Starting with a nominal sentence in the place of a verbal sentence"* occurred in the media, religious, linguistic, historical and technological texts (1 time each). For example, in the religious text the term 'Muslim historians say that …' was translated into مؤرخون مسلمون نقول '*Moarrekhoon Muslimoon naqool* ..' which means 'Muslim historian we say ..' It is supposed to be translated as يقول مؤرخون مسلمون '*Yaqoolu Moarrekhoon Muslimoon* ..'. The flexibility of Arabic to start with either the subject or the verb is based on the meaning. The nominal sentence starts with the subject to indicate stability while the verbal sentence starts with the verb to indicate actions at certain time. Another example is found in the linguistic text where Google Translate used the nominal sentence in the place of the verbal sentence. It translated the terms 'Linguists use the term varieties' as اللغويون استخدام مصطلح الاصناف '*Allughaweyoon istekhdam mustalah al-asnaaf'*. The meaning is slightly altered to 'The linguists' usage of the term varieties'. The verb 'use' was replaced with the noun 'usage'. The accurate translation of the phrase above into Arabic is: يستخدم اللغويون مصطلح الاصناف '*Yastakhdem Allughaweyoon mustalah al-asnaaf''*. The most candidate reason for such an error is that Google Translate is mainly programmed to follow the English sentence structure (i.e. subject + verb), but Arabic, as mentioned before, has the flexibility to start with either the subject or the verb based on the intended meaning.

   "Violating the whole phrase structure" had the highest occurrence within the category. It reached its peak in the linguistic and literary texts (9 and 7 times respectively). It appeared in the forms of putting adjectives before nouns, or putting modifiers before the modified terms. In Arabic, the adjective should be placed after the noun (which is different from English). For example, in the linguistic text, we read أكبر ثقافة المجتمع '*akbar thaqafatul mujtamaa'* to mean ثقافة المجتمع الكبرى '*thaqafatul mujtamaa al-kubra*' (English: The larger culture of the community). Again, such an error has led to an inflectional error in reference where the word اكبر '*akbar*' describes masculine noun while the term culture in Arabic is feminine noun that should be described by the adjective كبرى '*kubra*'. Likewise, in the philosophical text, the program translated the term 'logical fact' as منطقي الحقائق '*manteqi al haqaa'eq*' while the correct Arabic order is الحقائق المنطقية '*al haqaa'eq al manteqiah*'. The placing of the adjective before the noun has led to an inflectional error represented in using a masculine adjective منطقي '*manteqi*' to describe a plural feminine noun الحقائق '*al haqaa'eq*'. In another sentence, Google translated the adjective phrase 'circular argument' into حجة دائري '*hujjat daa'eri*', with the same syntactic and inflectional error. The word حجة '*hujjat*' is feminine while the word دائري '*daaeri*' is masculine. The possible interpretation of such an error is that Google follows the English adjective phrase structure where the adjective is put before the noun, with no difference between masculine or feminine nouns. In Arabic, the adjective must be placed after the noun and it should follow the case of the noun regarding masculinity and femininity, and singularity and plurality.

Grammatical errors

As for grammatical errors, *omitting functional morphemes* had the highest frequency in all categories (54 times). Errors of this type were found mostly in the educational text (13 times)

and the religious text (11 times). For instance, the phrase "Summative and formative assessments are often referred to in a learning context" was translated as وغالبا ما يشار التقييم التلخيصي والتكويني في سياق التعلم "*wa ghaliban ma yushaar at-taqyeem at-talkheesi wat-takweeni fi seyaq at-ta'allum*". The preposition الى "*ila*" which means "to" was not translated, although it is a necessary component of the Arabic prepositional phrase. The Arabic translation should be read in as: وغالبا ما يشار <u>الى</u> التقييم التلخيصي والتكويني في سياق التعلم "*wa ghaliban ma yushar <u>ila</u> at-taqyeem at-talkheesi wat-takweeni fi seyaq at-ta'allum*". Similarly, the term "similar to summative assessment" was translated into مشابه لتقييم تلخيصي "*mushabeh li taqyeem talkheesi*" where the article "the" "*al*" is omitted from the word تقييم *taqyeem* and the word تلخيصي *talkheesi*. Another example is taken from the religious text where the term "pertaining to Islamic jurisprudence" was translated into المتعلقة الفقه الاسلامي "*al muta'lleqatu al fiqh al Islami*" omitting the preposition "*bi*" that should precede the term "*al fiqh*". In the historical text, the conjunction و "*wa*" was omitted when translating the term "and the arrival of millions of immigrant workers". Such a term was translated as وصول الملايين من العمال المهاجرين "*wosool al malayeen men al ummal al muhajereen*" without the conjunction "*wa*" before the word "*wosool*".

    'Using wrong form of the word' counted 41 times (23.6% of the total grammatical errors) revealing the inadequacy of the Arabic grammatical functions which Google Translate uses. Such errors were mostly found in the technological text and manifested in the form of using the wrong form of the five verbs, the five nouns, and the inflections of nouns and verbs. For example, the term '*and its documentation*' in the technological text was translated into وثائقها '*wathaequha*' while the correct translation is وتوثيقاتها '*wa tawthiqatuha*'. In another sentence, the program used the verb صاغ '*saagha*' in the active voice to mean 'was coined' instead of صيغ '*seegha*' in the passive voice. A third example is also found in the technological text, where the phrase 'another piece' was translated into آخر قطعة '*aakhar qeta'a*', while it is supposed to be قطعة اخرى '*qeta'tun ukhra*'. In the media text, the term 'known for its global' was translated into تشتهر العالمية '*tashtaher al alameyah*' instead of تشتهر بعالميتها '*tashtaher biaalameyateha*'. In the geographical text, we read المتقاربة، المتباعدة او تحويل "*almutaqaarebah , al mutabaedah, awo tahweel*" as an equivalent translation of the English phrase "convergent, divergent, or transform". The accurate Arabic translation of such a phrase is المتقاربة، اوالمتباعدة او المتحولة "*almutaqarebah, aw al mutabaedah, aw almutahawwelah*". In the media text, the phrase للحد من تجاوز الصلاحيات تتركز في الحكومات '*llilhad men tajawoz assalahyat tatarakkaz fi al hakoomat*' was used as equivalent to the English phrase 'to limit the overreach of powers concentrated in governments'. The error is in the use of the verb تتركز '*tatarakkaz*' in the place of the noun المتركزة '*al mutarakkizah*'. It is worth mentioning that both philosophical and literary texts contained the lowest number of these errors (1 error each).

    'Using wrong prepositions, articles, and particles' came in the third place in grammatical errors, peaking in the technological text (7 times). For example, the English phrase: "concerned with the operation" was translated as المعنية **مع** تشغيل "*al ma'neyah ma'a tashgheel*". Google Translate used the preposition مع "*ma'a*" while the correct preposition is بـ "*bi*"; so the correct translation of such a phrase should be المعنية يتشغيل "*al ma'neyah be tashgheel*". Another example from the technological text is the translation of the phrase "binary values signifying processor instructions" into القيم الثنائية مما يدل على تعليمات المعالج "*al qeyam al thunaeyah mimma yadullu ala ta'leemat al mua'lej*". The dual preposition مما "*mimma*" here is inappropriate; the correct term should be the relative pronoun التي "*allati*". The Arabic translation of such a phrase should be read as القيم الثنائية التي تدل على تعليمات المعالج "*al qeyam al thunaeyah <u>allati</u> tadullu ala ta'leemat al mua'lej*". A third example can be taken from the religious text, where the term "and clarification

of" was translated into التوضيحات لو "*lou at-tawdheehat*". The use of the particle لو "*lou*" here is wrong; the correct particle should be و "*wa*" which means "and".

'Using a noun in place of a verb' was found 13 times in the translated texts (Table 5). In the physical text, for example, the word 'results' in the phrase "The understanding and use of acoustics *results* in better concert", was translated as النتائج "*annataaej*" which is equivalent to the noun 'the results'. The possible reason for such an error is the confusion, which Google Translate has in translating the inflectional morpheme '**s**' attached to the word 'results'. It is treated as the plural '**s**' rather than the '**s**' of the third person singular. Another example is found in the literary text where the term "American novelists *were expanding* fiction's social spectrum". The past progressive 'were expanding' was translated as توسيع "*towseea*'", the Arabic noun for وسع "*wassa'a*". A possible reason for that is the translation of each component of the past progressive alone. The program translated the phrase as وكان الروائيون الامريكي توسيع طيف الخيال" *wa kana ar-riwaaeyoon al ameriki towseea' taif al khayal*".

A serious error related to grammar is the use of the definite article before the modified and the modifier nouns in genitive cases. In Arabic, the definite article is attached to the modifier only. For example, in the religious text the translation of "the science of hadith" came as العلم الحديث "*al elm al hadeeth*" where the definite article was added to both modified and modifier nouns. Another example is found in the physical text in the translation of the term "Applied physicists". It was translated as للعلماء الفيزياء "*lilulama al fizeyaa*" where the definite article was used before the two nouns in such a genitive case.

Semantic errors

Semantic errors counted for 37.4% of the total errors. The highest item in such a category was no.15 "adding an unnecessary word, preposition, or article" (51 times). All the texts contained such an error with different rates. The media text contained the highest number of these errors (10 errors). In the religious text, for example, the addition of the definite article 'ال' to the above mentioned phrase "the science of hadith" results in different meaning, (i.e. *the modern science*) instead of the science of hadith. Also in the media text, the definite article was added to the word "*nashaat*" in الصحافة هي النشاط او منتج "*assahafatu heya an-nashaat aw montaj*", while it is supposed to be used without any article as it is indefinite, and should parallelize with other words in the sentence. The meaning of such a phrase became odd. Another example taken from the media text is in the translation of the term "the news media is government-controlled", where the relative pronoun التي 'allati' was used in the translation as وسائل الاعلام التي تسيطر عليها الحكومات "*wasaaellu e'elaam allati tusaiteru aliha al hokoomah*". The accurate translation is وسائل الاعلام تسيطر عليها الحكومات "*wasaaellu e'elaam tusaiteru aliha al hokoomah*" without the relative pronoun.

'Using a wrong meaning of the English homonyms' was also of high frequency in the texts being analyzed (37 times). Such an error happened more in the religious text, then in the philosophical and technological texts (Table 6). For example, in the religious text, the word 'in respect to' was translated as في احترام ل "*fi ehteraam li*" regardless of the context, while the correct meaning for it should be فيما يتعلق بـ "*fima yata'allaq bi*". Another example taken from the philosophical text where the verb 'states' with the third person singular 's' was translated into دول "*dowal*" 'Eng.= countries" rather than ينص "*yanuss*". In the technological text, the term 'code' was translated as قانون "qanoon" (law) while the equivalent word for that in Arabic is رمز "ramz" (symbol).

'Using wrong reference and relative pronoun' peaked in the educational text. For example, the term 'formative in nature' was translated as التكوينية في الطبيعة" *at-takweeneyah fi at-tabeea'h*" while the correct one should be التكويني بطبيعته" *takweeney bi tabeea'tihi*".

Google Translate also made another semantic error. That is 'using terms that convey very different meaning'. In the historical text, we read وسط مبدع "*wasat mubdea*" as an equivalent translation of the English phrase "central iconic event". The correct translation should be الحدث المميز "*al hadath al mumayaz*". The program also confused the word *'statics'* in the physical text with the word '*statistics*' and translated it into احصائيات "*ehsaaeyat*".

'Using words of ambiguous meaning' was another semantic error made by Google Translate in two texts: the philosophical and literary texts. In the philosophical text, the program translated the phrase 'by infinite regression' into من قبل الانحدار النهائي "*men qebal al enhidaar la nehaaeei*" while the correct translation is من حيث التراجع "*men haith at-taraju' al la nehaaeei*". Another example from the literary text is the translation of the verb 'scrutinized' into تمحيص "*tamhees*" while the correct equivalent verb in Arabic is فحصت "*fahasat*" or اختبرت "*ekhtabarat*".

It is worth mentioning that one of the problems which Google Translate faces in translation from English into Arabic is the lack of ability to find specialized vocabulary related to certain disciplines. For example, the program could not find the Arabic equivalent terms to the English terms *infinitism*, *foundationalism*, and *coherntism* in the philosophical text; and *deixis*, *sociolinguist*, and *ethnolinguist* in the linguistic text. The equivalent terms for these words are: عالم اللغة العرقي ,علم اللغة الاجتماعي ,السياقيات ,التماسكية ,القواعدية ,النهائية respectively. On the other hand, there were 4 sentences that had no errors. Such sentences were in the media, philosophical, literary, and educational texts.

Conclusions and Recommendations

The aim of the present study was to detect the language errors of Google Translate in translating texts from English into Arabic as an example of machine translation programs. The findings revealed three types of language errors: syntactic, grammatical, and semantic. Although Google Translate continues to improve the quality of machine translation, such findings support the claim of Aiken et al (2011b, p. 5) that the accuracy of Google Translate is affected by the type of pairs in which translation takes place and "translation comprehension still far from perfect because Google Translate's accuracy varies with sentence and vocabulary complexity and by language." Even with more advanced programs, machine translation is still in need for enormous effort to improve them, and, as Korosec (2011) says, translators need to remain aware of its limitations. The following are some recommendations and suggestions that may help improve future versions of Google Translate.

- It is recommended to get help from Arabic language experts, primarily those who are specialized in Arabic grammar, semantics, morphology, and syntax.
- Program developers may think about using Arabic vowel points (*Harakaat*) to help eliminate grammar errors.
- Contextual cohesive and coherent devices may play a vital role in improving the translation quality.
- Creating a free link between Google Translate users and administrators to benefit from updated suggestions, especially from users who are qualified in computer programming.

The question whether machine translation would replace human translation was and is still one of the primary concerns of research in machine translation. Researchers, in this regard, are between fear and confidence. Some look at it as a real threat to human translators; others are doubtful and base their doubt on the terrible errors committed by machine translation. Korosec (2011), for example, states that the current machine translation is nowhere near replacing the human translator. She doesn't deny, however, the role of machine translation in facilitating human translators' work and improving efficiency. Butler (2011, p. 9) negatively argues that machine translation will not substitute human translator. He bases his argument against machine translation on the nature of the language and the relationship between human, culture, and language. He states, "Humans are somewhat illogical beings and language is an adaptable, ever-changing, living concept that reflects the human psyche which may never be entirely captured in its essence by a machine." For machine translation tools to be used successfully, he suggests that the input needs to be simplified to a level where language losses its luster and beauty, and the message itself turns into an unmemorable static expression, dummied down to accommodate the limitations of machine language converters.

Supporters of machine translation, especially Google Translate clients, may argue that machine translation saves time and effort. No one can refute such an argument, but the question is whether the time and the effort spent in editing the output of machine translation is of less value! To establish a counter argument on scientific facts, a future research is needed to compare the time and the effort spent in translation texts by professional human translators to the time and effort spent in editing the output of machine translation, particularly in the case of Arabic language.

The present study is a knob in the range of machine translation research in general and Google Translate research in particular. It tackles language errors in translating from English into Arabic, which represents one face of the coin. More research is needed to show the other face (i.e. from Arabic into English). Previous studies in this regard were conducted at the level of word, phrase, and simple sentences, which necessitates future research to deal with text translation from Arabic into English. More focused research on every type of language errors by itself may set up an important database for Google Translate program developers. Further research is also needed to address contrastive linguistics issues, comparing Google Translate performance in translating the texts into other language. Besides, a research that investigates the reasons behind the limitations of Google Translate will contribute significantly to the program improvement. Google Translate Toolkit that was launched in 2009 was a remarkable addition to the program. However, it should not be the final step in such endeavor to improve the quality of translation.

About the Author:
Dr. Ahmad M. Al-Samawi is an Associate Professor of Applied Linguistics at Al Ain University of Science and Technology, UAE and Sana'a University, Yemen.

References

Aiken, M., & Vanjani, M. (2009). Polyglot: a multilingual group support system. Issues in Information Systems, 2, 101-106.

Aiken, M., Ghosh, K., Wee, J., & Vanjani, M. (2009a). An evaluation of the accuracy of online translation systems. Communications of the IIMA, 9(4), 67-84.

Aiken, M., Park, M., Simmons L., & Lindblom, T. (2009b). Automatic translation in multilingual electronic meetings. Translation Journal, 13(3). Retrieved from http://translationjournal.net/journal/49meeting.htm

Al-Kabi, M., Hailat, T., Al-Shawakfa, E., & Alsmadi, I. (2013). Evaluating English to Arabic machine translation using BLEU. (IJACSA) International Journal of Advanced Computer Science and Applications, 4(1), 66-73. Retrieved from www.ijacsa.thesai.org

Al-Samawi, A. (2012). The changing perception of literacy. Sana'a University Journal of Education and Psychological Sciences, 9(2), 1-28

Al-Samawi, A. (2013). Analysis of language errors in promotional circulars and their possible causes. European Journal of Social Sciences, 40(4), 649-668

ATA Software Technology Ltd. (2013). News-ata software. Available at http://www.atasoft.com/documents/33.html

Atlas Dictionaries. (2013). About us. Available at http://www.atlasdictionary.co/About_Us.aspx?key=Establishment

Balk, E. M., Chung, M., Chen, M. L., Trikalinos, T. A. & Kong, W. C. L. (2013). Assessing the accuracy of Google Translate to allow data extraction from trials published in non-English languages [Internet]. In Rockville (MD): Agency for Healthcare Research and Quality (US); 2013 Jan. Report No.: 12(13)-EHC145-EF.AHRQ Methods for Effective Health Care.

Berner, S. (2003). Lost in translation: cross-lingual communication, and virtual academic communities. Paper presented at the 5th Annual Conference on World Wide Web Applications, 10-12 September 2003, Durban, South Africa. Available at: http://www.samberner.com/documents/KM/lost.pdf

Brkic, M., Mikulic, B. B. & Matetic, M. (2012). Can we beat Google Translate? Information Technology Interfaces (ITI), Proceedings of the ITI 2012 34th International Conference on 25-28 June 2012 at Cavtat, Dubrovnik, pp381 – 386

Butler, F. (2011). Machine versus human: Will Google Translate replace professional translators? A paper presented at George Mason University. http://mason.gmu.edu/~fbutler2/IT%20103-005%20Research%20Paper%20Butler.pdf

Chin, J. (2013a). Introducing our new input tool. The official Google Translate Blog. Posted on February 6, 2013. http://googletranslate.blogspot.ae/#googtrans/en/en

Chin, J. (2013b). Get your YouTube video captions professionally translated into 36 languages. The official Google Translate Blog. Posted on February 20, 2013. http://googletranslate.blogspot.ae/#googtrans/en/en

Corder, S.P. (1967). The significance of learner's errors. IRAL 5, 161-170

Corder, S.P. (1981). Error analysis and interlanguage. Oxford: Oxford University Press

Crystal, D. (2008). A Dictionary of Linguistics and Phonetics (6th ed). London: Blackwell Publishing.

Dhore, M. L., Dixit, S. K. & Sonwalker, T. D. (2012). Hindi to English machine transliteration of named entities using conditional random fields. International Journal of Computer Applications, 48(June), pp. 31-37

Dictionaries and Encyclopedia. (2013). Babylon (Software). The Free Encyclopedia (Wikipedia). http://en.wikipedia.org/wiki/Babylon_(software)

Diplo foundation. (2011) What's machine translation? Retrieve from: www.diplomacy.edu/language/Translation/machine.htm (March, 2011)

ElShiekh, A. A. (2012). Google Translate service: transfer of meaning, distortion or simply a new creation? An investigation into the translation process & problems at Google. English Language and Literature Studies, 2(1) 56-68. http://dx.doi.org/10.5539/ells.v2n1p56

Farrús, M., Ruiz Costa-Jussà, M., Mariño A. J. & Rodríguez F. J. (2010). Linguistic-based evaluation criteria to identify statistical machine translation errors. A paper presented at the Annual Conference of the European Association for Machine Translation. 14th Annual Conference of the European Association for Machine Translation. Saint-Raphaël: 2010, p. 167-173. http://hdl.handle.net/2117/7492

Google Translate. (2013). About Google Translate. http://translate.google.com/about/

Hujair, J. (2012). Atta'reeb fi Google "Arabizing in Google" (An interview with Dr. Fayeq Oweis) Aa'lam At-Taqneyah. Available at: http://www.tech-wd.com/wd/2012/03/18/arabic-localization-in-google/

Hutchins, W. J. & Somers, H. L. (1992). An introduction to Machine Translation. UK: Academic Press Limited

Jamilah, N. (2012). Semantic translation errors as a result of Google Translate (a case of errors of translation on homonymous and polysemous words in Bahasa Indonesia). Diglossia 4(1) Retrieved from http://www.journal.unipdu.ac.id/index.php/diglosia/issue/view/27.pdf

Jiang, M. (2013a). The world's languages in your pocket (No Internet Required). The official Google Translate Blog. Posted on March 27, 2013. At: http://googletranslate.blogspot.ae/#googtrans/en/en

Jiang, M. (2013b). Take your Phrasebook on the go—and translate by camera in 16 more languages. The official Google Translate Blog. Posted on May 8, 2013. http://googletranslate.blogspot.ae/#googtrans/en/en

Kelman, S. (2013). More than 70 of the world's languages in the blink of an eye. The official Google Translate Blog. Posted on Wednesday, May 08, 2013. http://googletranslate.blogspot.ae/#googtrans/en/en

Korosec, M. K. (2011). The Internet, Google Translate and Google Translator Toolkit - nuisance or necessity in translator training? Retrieved from: Auditorium Du Cnrs, Paris le 3 et 4 mars 2011.

Larsen-Freeman, D. & Long, M. (1991). An introduction to second language acquisition research. UK: Longman

Och, F. (2006). Statistical machine translation live. Google Research Blog, Posted on April, 2006 at: http://googleresearch.blogspot.ae/2006/04/statistical-machine-translation-live.html

Resnik, P., Buzek, O., Hu, C., Kronrod, Y., Quinn, A. & Bederson, B. B. (2010). Improving translation via targeted paraphrasing. Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pp. 127–137, MIT, Massachusetts, USA, 9-11 October 2010

Schwartz, B. (2007). Google Translate drops systran for home brewed translation. http://searchengineland.com/google-translate-drops-systran-for-home-brewed-translation-12502

Selijan, S., Brkic, M. & Kucis, V. (2011). Evaluation of free online machine translations for Croatian-English and English-Croatian language pairs. Paper presented at 3rd International Conference of The future of Information Science (INFuture) in Zagreb, 9-11 November 2011. pp. 331-340

Stemler, S. (2001). An overview of content analysis. Practical Assessment, Research & Evaluation, 7(17). Retrieved on May 24, 2013. http://PAREonline.net/getvn.asp?v=7&n=17

Wikipedia. (2013). The Free Encyclopedia. http://en.wikipedia.org/wiki/Encyclopedia
Zervaki, T., 2002. Online free translation services. Proceedings of the 24th International Conference on translating and the Computer. London, UK.

Appendix (A)
Original Texts in English and their Translation into Arabic by Google Translate.
Text 1: Journalism

1. Journalism is the activity, or product, of journalists or others engaged in the preparation of written, visual, or audio material intended for dissemination through public media with reference to factual, ongoing events of public concern.

2. It is intended to inform society about itself and to make public events that would otherwise remain private.

3. In modern society, news media are the chief purveyor of information and opinion about public affairs.

4. Journalism, however, is not to be confused with the news media or the news itself.

5. In some nations, the news media is government-controlled and not an independent body that operates within journalistic frameworks.

6. In democratic societies, access to information can play a key role in a system of checks and balances designed to limit the overreach of powers concentrated in governments, businesses and other entities and individuals.

7. Access to verifiable information gathered by independent media sources adhering to journalistic standards can also provide ordinary citizens with the tools they need to participate in the political process.

8. The role and status of journalism, along with mass media, have undergone profound changes resulting from the publication of news on the Internet.

9. This has created a shift away from print media consumption as people increasingly consume news on e-readers, smartphones, and other electronic devices, challenging news organizations to fully monetize digital news.

10. Notably, in the American media landscape, newsrooms have reduced their staff and coverage as traditional media channels such as television grapple with declining audiences; for instance, at CNN, once known for its global, in-depth coverage, produced story packages were cut nearly in half from 2007 to 2012.

Source: Wikipedia (2013) *The Free Encyclopedia*.

النص الاول- الصحافة
1. الصحافة هي النشاط، أو منتج، من الصحفيين أو غيرهم من العاملين في إعداد كتابي، المواد البصرية، أو السمعية المعدة للنشر من خلال وسائل الإعلام العامة مع الإشارة إلى واقعية، والأحداث الجارية ذات الاهتمام العام.
2. فإنه يهدف إلى إعلام المجتمع عن نفسه وجعل المناسبات العامة التي من شأنها أن تبقى على خلاف ذلك خاصة.
3. في المجتمع الحديث، وسائل الإعلام هي المزود الرئيسي للمعلومات والرأي حول الشؤون العامة.
4. الصحافة، ومع ذلك، لا ينبغي الخلط بينه وبين وسائل الإعلام أو الأخبار نفسها.
5. في بعض الدول، وسائل الإعلام التي تسيطر عليها الحكومة هو وليس هيئة مستقلة تعمل ضمن الأطر الصحفية.
6. في المجتمعات الديمقراطية، والوصول إلى المعلومات يمكن أن تلعب دورا رئيسيا في نظام من الضوابط والتوازنات مصممة للحد من تجاوز الصلاحيات تتركز في الحكومات والشركات والكيانات والأفراد الآخرين.
7. الوصول إلى معلومات يمكن التحقق منها التي جمعتها مصادر إعلامية مستقلة تتقيد بالمعايير الصحفية يمكن أن توفر أيضا المواطنين العاديين مع الأدوات التي يحتاجونها للمشاركة في العملية السياسية.
8. دور ومكانة الصحافة، جنبا إلى جنب مع وسائل الإعلام، فقد شهد تغيرات عميقة الناتجة عن نشر أخبار على شبكة الإنترنت.
9. وقد خلق هذا التحول بعيدا عن استهلاك وسائل الإعلام المطبوعة والناس تستهلك بشكل متزايد الاخبارية على القراءة الإلكترونية، والهواتف الذكية، والأجهزة الإلكترونية الأخرى، مما يشكل تحديا وكالات الأنباء لنقد بالكامل الأخبار الرقمية.
10. والجدير بالذكر، في المشهد الإعلامي الأمريكي، خفضت غرف الأخبار موظفيها وتغطية وسائل الإعلام التقليدية مثل التلفزيون التعامل مع انخفاض الجماهير، على سبيل المثال، في CNN، التي كانت تشتهر العالمية، تغطيتها في العمق، أنتجت قطعت حزم قصة تقريبا في النصف 2007- 2012.

Text 2: Hadith in Islam

1. The overwhelming majority of Muslims consider hadith to be essential supplements to and clarifications of the Quran, Islam's holy book, as well as in clarifying issues pertaining to Islamic jurisprudence.

2. Ibn al-Salah, a hadith specialist, described the relationship between hadith and other aspect of the religion by saying: "It is the science most pervasive in respect to the other sciences in their various branches, in particular to jurisprudence being the most important of them."

3. "The intended meaning of 'other sciences' here are those pertaining to religion," explains Ibn Hajar al-Asqalani, "Quranic exegesis, hadith, and jurisprudence.

4. The science of hadith became the most pervasive due to the need displayed by each of these three sciences.

5. The need hadith has of its science is apparent.

6. As for Quranic exegesis, then the preferred manner of explaining the speech of God is by means of what has been accepted as a statement of Muhammad.

7. The one looking to this is in need of distinguishing the acceptable from the unacceptable.

8. Regarding jurisprudence, then the jurist is in need of citing as an evidence the acceptable to the exception of the later, something only possible utilizing the science of hadith.

9. Traditions of the life of Muhammad and the early history of Islam were passed down mostly orally for more than a hundred years after Muhammad's death in AD 632.

10. Muslim historians say that Caliph Uthman ibn Affan (the third khalifa (caliph) of the Rashidun Empire, or third successor of Muhammad, who had formerly been Muhammad's secretary), is generally believed to urge Muslims to record the hadith just as Muhammad suggested to some of his followers to write down his words and actions.

Source: Wikipedia (2013) *The Free Encyclopedia*.

النص الثاني: الحديث والسيرة عند المسلمين

1. الأغلبية الساحقة من المسلمين يعتبرون الحديث لتكون ملاحق أساسية لوالتوضيحات من القرآن الكريم، كتاب الإسلام المقدس، وكذلك في توضيح القضايا المتعلقة الفقه الإسلامي.

2. ووصف بن صلاح، وهو متخصص الحديث، والعلاقة بين الحديث وجانب آخر من جوانب الدين قائلا: "هذا هو العلم الأكثر انتشارا في احترام لغيرها من العلوم في فروعها المختلفة، وعلى وجه الخصوص إلى فقه كونها أهم منهم ".

3. "والمعنى المقصود من" العلوم الأخرى "هنا هي تلك التي تتعلق بالدين"، ويوضح ابن حجر آل Asqalani، "تفسير القرآن الكريم، الحديث، والفقه.

4. وأصبح العلم الحديث الأكثر انتشارا بسبب الحاجة إلى عرض من كل من هذه 3 العلوم.

5. الحاجة لديه من العلم الحديث لها هو واضح.

6. أما بالنسبة لتفسير القرآن الكريم، ثم الطريقة المفضلة لشرح كلمة الله هي من خلال ما تم قبوله بيان من محمد.

7. واحد يبحث في هذا هو في حاجة على تمييز المقبول من غير المقبول.

8. وفيما يتعلق فقه، ثم الفقيه هو في حاجة إلى ذكر على أدلة مقبولة لدى باستثناء في وقت لاحق، وهو أمر ممكن فقط الاستفادة من العلم الحديث.

9. صدرت تقاليد حياة محمد وتاريخ الإسلام في وقت مبكر من أسفل شفويا في الغالب لأكثر من مائة سنة بعد وفاة النبي محمد في 632 للميلاد.

10. مؤرخون مسلمون نقول ان يعتقد عموما الخليفة عثمان بن عفان (خليفة الثالث (الخليفة) للإمبراطورية الراشدين، أو خليفة الثالثة لمحمد، الذي كان سابقا أمين محمد)، لحث المسلمين على تسجيل الحديث مثلما محمد اقترح على بعض من أتباعه لكتابة كلماته وتصرفاته.

Text 3: Epistemology

1. Epistemology is concerned with the nature and scope of knowledge, such as the relationships between truth, belief, and theories of justification1.

2. Skepticism is the position that questions the possibility of justifying any truth.

3. Münchhausen Trilemma states that the three options to soundly prove any truth are not satisfactory.

4. One is the regress argument where, by infinite regression, each proof requires a further proof.

5. Infinitism claims that the chain can go forever.

6. Another is foundationalism, where justification eventually rests on unproven basic beliefs or axioms.

7. Logical atomism holds there are logical "facts" (or "atoms") that cannot be broken down any further.

8. The other method of justification involves the circular argument, in which theory and proof support each other.

9. Coherentism claims a belief is justified if it coheres with the larger belief system.

10. More specifically, the coherence theory of truth states what is true is that which coheres with some specified set of propositions.

Source: Wikipedia (2013) *The Free Encyclopedia*.

النص الثالث: نظرية المعرفة

<div dir="rtl">

1. ويساور نظرية المعرفة مع طبيعة ونطاق المعرفة، مثل العلاقات بين الحقيقة والاعتقاد، ونظريات التبرير.
2. الشك هو الموقف الذي يشكك في إمكانية تبرير أي حقيقة.
3. Münchhausen Trilemma تنص على أن الخيارات الثلاثة لإثبات أي حقيقة على نحو سليم ليست مرضية.
4. واحدة هي حجة التراجع حيث، من قبل الانحدار لانهائي، كل برهان يتطلب دليلا آخر.
5. Infinitism يدعي أن سلسلة يمكن أن تذهب إلى الأبد.
6. آخر هو foundationalism، حيث تقع في نهاية المطاف على تبرير المعتقدات الأساسية غير مثبتة أو البديهيات.
7. مذهب الذرية منطقي يحمل هناك منطقي "الحقائق" (أو "الذرات") التي لا يمكن تقسيمها إلى أبعد من ذلك.
8. طريقة أخرى من مبرر ينطوي على حجة دائري، في النظرية والتي برهان يدعم كل منهما الآخر.
9. Coherentism يدعي له ما يبرره إذا كان الاعتقاد يتماسك مع نظام أكبر المعتقد.
10. وبشكل أكثر تحديدا، فإن نظرية الترابط بين الدول حقيقة ما هو صحيح هو الذي يتماسك مع بعض مجموعة محددة من المقترحات.

</div>

Text 4: Language and Culture

1. Languages, understood as the particular set of speech norms of a particular community, are also a part of the larger culture of the community that speak them.
2. Humans use language as a way of signalling identity with one cultural group and difference from others.
3. Even among speakers of one language several different ways of using the language exist, and each is used to signal affiliation with particular subgroups within a larger culture.
4. Linguists and anthropologists, particularly sociolinguists, ethnolinguists and linguistic anthropologists have specialized in studying how ways of speaking vary between speech communities.
5. A community's ways of using language is a part of the community's culture, just as other shared practices are, it is way of displaying group identity.
6. Ways of speaking function not only to facilitate communication, but also to identify the social position of the speaker.
7. Linguists use the term varieties, a term that encompasses geographically or socioculturally defined dialects as well as the jargons or styles of subcultures, to refer to the different ways of speaking a language.
8. Linguistic anthropologists and sociologists of language define communicative style as the ways that language is used and understood within a particular culture.
9. Languages do not differ only in pronunciation, vocabulary or grammar, but also through having different "cultures of speaking".
10. Some cultures for example have elaborate systems of "social deixis", systems of signalling social distance through linguistic means.

Source: Wikipedia (2013) *The Free Encyclopedia*.

<div dir="rtl">

النص الرابع: اللغة والثقافة

1. اللغات، والتي يفهم منها مجموعة معينة من قواعد الكلام في مجتمع معين، هي أيضا جزء من أكبر ثقافة المجتمع التي تتحدث بها.
2. البشر يستخدمون اللغة كوسيلة للإشارة الهوية الثقافية مع مجموعة واحدة والاختلاف عن الآخرين.
3. حتى بين المتحدثين بلغة واحدة من طرق عديدة ومختلفة من استخدام لغة الوجود، ويستخدم للإشارة إلى كل الانتماءات مع مجموعات فرعية خاصة ضمن أكبر ثقافة.
4. وقد تخصصت اللغويين وعلماء الانثروبولوجيا، sociolinguists خاصة، ethnolinguists وعلماء الانثروبولوجيا اللغوية في دراسة كيفية سبل يتحدث تختلف بين المجتمعات خطاب.
5. الطرق في أي مجتمع من استخدام اللغة هو جزء من ثقافة المجتمع، تماما كما غيرها من الممارسات المشتركة هي، هو طريقة لعرض هوية المجموعة.
6. طرق وظيفة الناطق، ليس فقط لتسهيل عملية التواصل، ولكن أيضا لتحديد الوضع الاجتماعي للمتكلم.
7. اللغويون استخدام مصطلح الأصناف، وهو مصطلح يشمل اللهجات جغرافيا أو اجتماعيا وثقافيا محددة، فضلا عن المفردات التخصصية أو أنماط من الثقافات الفرعية، للإشارة إلى طرق مختلفة ليتحدث لغة.
8. علماء الأنثروبولوجيا وعلماء الاجتماع لغوية للغة تحديد أسلوب التواصل والطرق التي يتم فيها استخدام اللغة وفهمها في إطار ثقافة معينة.
9. لغات لا تختلف فقط في المفردات، والنطق أو النحوي، ولكن أيضا من خلال وجود مختلف "ثقافات الكلام".
10. بعض الثقافات على سبيل المثال لديها نظم معقدة من "deixis الاجتماعية"، وأنظمة الإشارات المسافة الاجتماعية من خلال وسائل لغوية.

</div>

Text 5: Applied Physics

1. Applied physics is a general term for physics research which is intended for a particular use.
2. An applied physics curriculum usually contains a few classes in an applied discipline, like geology or electrical engineering.

3. It usually differs from engineering in that an applied physicist may not be designing something in particular, but rather is using physics or conducting physics research with the aim of developing new technologies or solving a problem.
4. The approach is similar to that of applied mathematics.
5. Applied physicists can also be interested in the use of physics for scientific research.
6. For instance, people working on accelerator physics might seek to build better particle detectors for research in theoretical physics.
7. Physics is used heavily in engineering.
8. For example, Statics, a subfield of mechanics, is used in the building of bridges and other structures.
9. The understanding and use of acoustics results in better concert halls; similarly, the use of optics creates better optical devices.
10. An understanding of physics makes for more realistic flight simulators, video games, and movies, and is often critical in forensic investigations.

Source: Wikipedia (2013) *The Free Encyclopedia*.

النص الخامس -الفيزياء التطبيقية

1. الفيزياء التطبيقية هو مصطلح عام لبحوث الفيزياء الذي يهدف لاستخدام معين.
2. منهج الفيزياء التطبيقية عادة ما تحتوي على فئات قليلة الانضباط في تطبيقها، مثل الجيولوجيا أو الهندسة الكهربائية.
3. وهو يختلف عادة من الهندسة في أن عالم الفيزياء التطبيقية قد لا يمكن تصميم شيء على وجه الخصوص، وإنما يستخدم الفيزياء أو إجراء بحوث الفيزياء وذلك بهدف تطوير تقنيات جديدة أو حل مشكلة.
4. هذا النهج هو مماثلة لتلك التي الرياضيات التطبيقية.
5. يمكن للعلماء الفيزياء التطبيقية تكون مهتمة ايضا في استخدام الفيزياء للبحث العلمي.
6. على سبيل المثال، قد يكون الناس الذين يعملون في الفيزياء مسرع نسعى لبناء أفضل للكشف عن الجسيمات للبحث في الفيزياء النظرية.
7. ويستخدم بشكل مكثف في الهندسة والفيزياء.
8. على سبيل المثال، يتم استخدام احصائيات، أحد فروع الميكانيكا، في بناء الجسور والمنشآت الأخرى.
9. فهم واستخدام النتائج الصوت في قاعات أفضل حفلة موسيقية، وبالمثل، فإن استخدام البصريات يخلق أفضل الأجهزة البصرية.
10. فهم الفيزياء يجعل لمحاكاة الطيران أكثر واقعية، وألعاب الفيديو، والأفلام، وغالبا ما تكون حاسمة في تحقيقات الطب الشرعي.

Text 6: American Literature in the 20th century

1. At the beginning of the 20th century, American novelists were expanding fiction's social spectrum to encompass both high and low life and sometimes connected to the naturalist school of realism.
2. In her stories and novels, Edith Wharton (1862–1937) scrutinized the upper-class, Eastern-seaboard society in which she had grown up.
3. One of her finest books, *The Age of Innocence*, centers on a man who chooses to marry a conventional, socially acceptable woman rather than a fascinating outsider.
4. At about the same time, Stephen Crane (1871–1900), best known for his Civil War novel *The Red Badge of Courage*, depicted the life of New York City prostitutes in *Maggie: A Girl of the Streets*.
5. And in *Sister Carrie*, Theodore Dreiser (1871–1945) portrayed a country girl who moves to Chicago and becomes a kept woman.
6. Hamlin Garland and Frank Norris wrote about the problems of American farmers and other social issues from a naturalist perspective.
7. More directly political writings discussed social issues and power of corporations.
8. Some like Edward Bellamy in *Looking Backward* outlined other possible political and social frameworks.
9. Upton Sinclair, most famous for his muck-raking novel *The Jungle*, advocated socialism.
10. Other political writers of the period included Edwin Markham, William Vaughn Moody.

Source: Wikipedia (2013) *The Free Encyclopedia*.

النص السادس: الادب الامريكي في القرن العشرين

1. في بداية القرن 20، وكان الروائيون الأمريكيون توسيع طيف الخيال الاجتماعي ليشمل الحياة على حد سواء المرتفعة والمنخفضة ومتصلة أحيانا إلى المدرسة الواقعية في الطبيعة.

2. في قصصها ورواياتها، تمحيص إديث وارتون (1862-1937) في الطبقة العليا، والشرقية، الساحل المجتمع الذي كانت قد كبروا.
3. واحد من أروع الكتب لها، عصر البراءة، ومراكز للرجل الذي يختار أن يتزوج من التقليدية، والمرأة مقبولة اجتماعيا، وليس من الخارج رائعة.
4. في الوقت نفسه تقريبا، يصور ستيفن كرين (1871-1900) الذي اشتهر لروايته الحرب الأهلية الشارة الحمراء من الشجاعة، حياة العاهرات في مدينة نيويورك في ماغي: وهي فتاة من الشوارع.
5. والأخت كاري، صورت ثيودور Dreiser (1871-1945) فتاة البلد الذي انتقل الى شيكاغو، وتصبح امرأة أبقيت.
6. كتب هاملين جارلاند وفرانك نوريس عن مشاكل المزارعين الأميركيين والقضايا الاجتماعية الأخرى من منظور طبيعي.
7. أكثر مباشرة وناقش الكتابات السياسية قضايا اجتماعية وقوة الشركات.
8. بعض إدوارد بيلامي في مثل يبحث المبينة متخلفة أخرى ممكن الأطر السياسية والاجتماعية.
9. أبتون سنكلير، الأكثر شهرة لالغابة له ارض خصبة، يخدش رواية، والاشتراكية تدعو.
10. شمل الكتاب السياسية الأخرى في فترة إدوين ماركهام، وليام فون مودي.

Text 7: Plate Tectonic

1.   The lithosphere is broken up into tectonic plates.
2.   On Earth, there are seven or eight major plates (depending on how they are defined) and many minor plates.
3.   Where plates meet, their relative motion determines the type of boundary: convergent, divergent, or transform.
4.   Earthquakes, volcanic activity, mountain-building, and oceanic trench formation occur along these plate boundaries.
5.   The lateral relative movement of the plates typically varies from zero to 100 mm annually.[2]
6.   Tectonic plates are composed of oceanic lithosphere and thicker continental lithosphere, each topped by its own kind of crust.
7.   Along convergent boundaries, subduction carries plates into the mantle; the material lost is roughly balanced by the formation of new (oceanic) crust along divergent margins by seafloor spreading.
8.   In this way, the total surface of the globe remains the same.
9.   This prediction of plate tectonics is also referred to as the conveyor belt principle.
10.   Earlier theories (that still have some supporters) proposed gradual shrinking (contraction) or gradual expansion of the globe.[3]

Source: Wikipedia (2013) *The Free Encyclopedia*.

النص السابع: الصفائح التكتونية

1. يتم تقسيم الغلاف الصخري تصل إلى الصفائح التكتونية.
2. على الأرض، وهناك سبعة أو ثمانية لوحات رئيسية (اعتمادا على كيفية تعريف هم) والعديد من لوحات طفيفة.
3. حيث يلتقي لوحات، ويحدد حركتها النسبية نوع من الحدود: المتقاربة، المتباعدة، أو تحويل
4. الزلازل ، النشاط البركاني، وبناء الجبال، والمحيطات تشكيل خندق تحدث على طول هذه الحدود لوحة.
5. الحركة النسبية الجانبي للوحات يختلف عادة من صفر إلى 100 ملم سنويا.[
6. وتتكون الصفائح التكتونية المحيطية من الغلاف الصخري والغلاف الصخري القاري سمكا، كل على رأسها نوع خاص بها من القشرة.
7. على طول الحدود المتقاربة، الاندساس يحمل لوحات في عباءة؛ المواد المفقودة متوازنة تقريبا من تشكيل جديد (المحيطية) القشرة على طول هوامش متباينة من قبل امتداد قاع البحر.
8. في هذه الطريقة، فإن المساحة الإجمالية من العالم لا يزال هو نفسه.
9. ويشار إلى هذا التوقع من الصفائح التكتونية أيضا على أنها مبدأ الحزام الناقل.
10. النظريات السابقة (التي لا تزال لديها بعض أنصار) اقترح تقلص تدريجي (انكماش) أو التوسع التدريجي في العالم.

Text 8: History of USA

1.   The 1860 election of Abraham Lincoln, who called for no more expansion of slavery, triggered a crisis as eleven slave states seceded to found the Confederate States of America in 1861.
2.   The bloody American Civil War (1861–65) redefined the nation and remains the central iconic event.
3.   The South was defeated and, in the Reconstruction era, the U.S. ended slavery, extended rights to African Americans, and readmitted secessionist states with loyal governments.
4.   The national government was much stronger, and it now had the explicit duty to protect individuals.
5.   Reconstruction was never completed by the US government and left the blacks in a world of Jim Crow political, social and economic inferiority.

6. The entire South remained poor while the North and West grew rapidly.
7. Thanks to an outburst of entrepreneurship in the North and the arrival of millions of immigrant workers from Europe, the U.S. became the leading industrialized power by 1900.
8. Disgust with corruption, waste, and traditional politics stimulated the Progressive movement, 1890s-1920s, which pushed for reform in industry and politics and put into the Constitution women's suffrage and Prohibition of alcohol (the latter repealed in 1933).
9. Initially neutral in World War I, the U.S. declared war on Germany in 1917, and funded the Allied victory.
10. The nation refused to follow President Woodrow Wilson's leadership and never joined the League of Nations.

Source: Wikipedia (2013) *The Free Encyclopedia*.

النص الثامن: تاريخ الولايات المتحدة
١. فجرت انتخابات 1860 من ابراهام لنكولن، الذي دعا الى المزيد من التوسع لا العبودية، وأزمة الولايات الرقيق كما انفصلت 11 لتأسيس الولايات الكونفدرالية الأميركية في عام 1861.
٢. إعادة تعريف دموية الحرب الأهلية الأمريكية (1861-1865) للأمة، ولا يزال هذا الحدث وسط مبدع.
٣. هزم الجنوب و، في عصر إعادة الإعمار، والولايات المتحدة انتهت العبودية وحقوق الممتدة إلى الأميركيين الأفارقة، والدول الانفصالية أعيدوا مع الحكومات الموالية.
٤. وكانت الحكومة الوطنية أقوى بكثير، وكان الآن واجب واضح لحماية الأفراد.
٥. لم يكتمل إعادة البناء من قبل حكومة الولايات المتحدة وترك السود في عالم جيم كرو السياسية، والدونية الاجتماعية والاقتصادية.
٦. ظلت جنوب كامل الفقيرة، بينما في الشمال والغرب نمت بسرعة.
٧. بفضل فورة من المشاريع في الشمال وصول الملايين من العمال المهاجرين من أوروبا، أصبحت الولايات المتحدة القوة الصناعية الكبرى بحلول عام 1900.
٨. حفز الاشمئزاز من الفساد والهدر والسياسة التقليدية الحركة التقدمية، 1890 s-1920s-، والتي دفعت من أجل الإصلاح في الصناعة والسياسة وضعت في الدستور حق التصويت للنساء، وتحريم الخمر (وهذا الأخير يلغى في 1933).
٩. أعلنت الولايات المتحدة محايدة في البداية في الحرب العالمية الأولى، الحرب على ألمانيا في عام 1917، ومولت انتصار الحلفاء.
١٠. ورفض الأمة لمتابعة قيادة الرئيس وودرو ويلسون وأبدا انضم إلى عصبة الأمم.

Text 9: Educational Assessment

1. Summative and formative assessments are often referred to in a learning context as *assessment of learning* and *assessment for learning* respectively.
2. Assessment of learning is generally summative in nature and intended to measure learning outcomes and report those outcomes to students, parents, and administrators.
3. Assessment of learning generally occurs at the conclusion of a class, course, semester, or academic year.
4. Assessment for learning is generally formative in nature and is used by teachers to consider approaches to teaching and next steps for individual learners and the class.
5. A common form of formative assessment is *diagnostic assessment*.
6. Diagnostic assessment measures a student's current knowledge and skills for the purpose of identifying a suitable program of learning.
7. *Self-assessment* is a form of diagnostic assessment which involves students assessing themselves.
8. *Forward-looking assessment* asks those being assessed to consider themselves in hypothetical future situations.
9. *Performance-based assessment* is similar to summative assessment, as it focuses on achievement.
10. It is often aligned with the standards-based education reform and outcomes-based education movement.

Source: Wikipedia (2013) *The Free Encyclopedia*.

النص التاسع: القياس التربوي
١. وغالبا ما يشار التقييم التلخيصي والتكويني لفي سياق التعلم وتقييم التعلم والتقييم للتعلم على التوالي.
٢. تقييم التعلم هو عموما تلخيصي في الطبيعة، وتهدف إلى قياس نتائج التعلم وتقديم تقرير عن تلك النتائج إلى الطلاب وأولياء الأمور، والإداريين.
٣. تقييم التعلم يحدث عادة في ختام فئة، فصل دراسي، بطبيعة الحال، أو العام الدراسي.

<div dir="rtl">

4. التقييم من أجل التعلم عموما التكوينية في الطبيعة، ويستخدم من قبل المعلمين للنظر في نهج التدريس والخطوات المقبلة للمتعلمين الفرد والطبقة.

5. شكلا شائعا من أشكال التقييم التكويني هو تقييم التشخيص.

6. التقييم التشخيصي يقيس معرفة الطالب الحالية والمهارات لغرض تحديد برنامج مناسب للتعلم.

7. التقييم الذاتي هو شكل من أشكال التقييم التشخيصي الذي ينطوي على طلاب تقييم أنفسهم.

8. تقييم تطلعي يسأل هؤلاء يجري تقييم للنظر في أنفسهم في حالات افتراضية في المستقبل.

9. على أساس الأداء تقدير مشابه لتقييم تلخيصي، كما أنها تركز على الإنجاز.

10. يتم محاذاة في كثير من الأحيان مع إصلاح التعليم المستندة إلى المعايير والنتائج المستندة إلى حركة التعليم.

</div>

Text 10: Software

1.   Software refers to one or more computer programs and data held in the storage of the computer.
2.   In other words, software is a set of *programs, procedures, algorithms* and its *documentation* concerned with the operation of a data processing system.
3.   Program software performs the function of the program it implements, either by directly providing instructions to the digital electronics or by serving as input to another piece of software.
4.   The term was coined to contrast to the term *hardware* (meaning physical devices).
5.   In contrast to hardware, software "cannot be touched".[1]
6.   Software is also sometimes used in a more narrow sense, meaning application software only.
7.   Sometimes the term includes data that has not traditionally been associated with computers, such as film, tapes, and records.[2]
8.   Computer software is so called to distinguish it from computer hardware, which encompasses the physical interconnections and devices required to store and execute (or run) the software.
9.   At the lowest level, executable code consists of machine language instructions specific to an individual processor.
10.   A machine language consists of groups of binary values signifying processor instructions that change the state of the computer from its preceding state.

Source: Wikipedia (2013) *The Free Encyclopedia*.

<div dir="rtl">

النص العاشر: التكنولوجيا

1. البرنامج يشير إلى واحد أو أكثر من برامج الكمبيوتر والبيانات التي عقدت في تخزين الكمبيوتر.

2. وبعبارة أخرى، البرنامج هو مجموعة من البرامج والإجراءات والخوارزميات وثائقها المعنية مع تشغيل نظام معالجة البيانات.

3. برنامج برنامج يؤدي وظيفة البرنامج التي ينفذها، إما مباشرة من خلال توفير تعليمات للإلكترونيات الرقمية أو من خلال خدمة كمدخل إلى آخر قطعة من البرمجيات.

4. وقد صاغ هذا المصطلح إلى النقيض من الأجهزة الأجل (بمعنى الأجهزة الفعلية.(

5. وعلى النقيض من الأجهزة والبرامج "لا يمكن لمسها". [1]

6. أيضا يتم استخدام البرمجيات في بعض الأحيان في أكثر بالمعنى الضيق، وهذا يعني تطبيق البرمجيات فقط.

7. أحيانا يشمل هذا المصطلح البيانات التي لم تقليديا ارتبطت مع أجهزة الكمبيوتر، مثل الأفلام والأشرطة والسجلات. [2]

8. ويسمى برامج الكمبيوتر من أجل تمييزه عن أجهزة الكمبيوتر، والذي يشمل الترابط المادي والأجهزة اللازمة لتخزين وتنفيذ (أو تشغيل) البرنامج.

9. عند أدنى مستوى، ويتألف قانون قابل للتنفيذ من تعليمات لغة الآلة محددة لمعالج الفردية.

10. يتكون لغة الآلة من مجموعات من القيم الثنائية مما يدل على تعليمات المعالج أن تغيير حالة جهاز الكمبيوتر من حالته السابقة.

</div>

Appendix (B): Summary of errors and their types in the examined texts

No	Error Category	Type of Error	Media	Religious	Philosophical	Linguistic	Physical	Literary	Geographic	Historical	Educational	Technological	Total
			Field of text and frequency of errors										
1	Syntactic Errors	Starting with a nominal sentence in the place of a verbal sentence.	1	1	0	1	0	0	0	1	0	1	5
2		Violating the whole phrase structure	4	3	5	9	5	7	5	6	3	3	50
3	Grammar Errors	Using wrong form of the word (plural, the five verbs, five nouns, nouns an d verbs inflections)	7	2	1	2	3	1	7	4	3	11	41
4		Violating subject-verb agreement (masculine and feminine; singular, dual, and plural; first, second, and third person)	0	2	1	3	2	4	4	4	1	1	22
5		Using a noun in place of a verb	1	1	0	2	3	3	2	1	0	0	13
6		Using a verb in place of a noun	0	0	0	1	0	0	0	0	0	0	1
7		Using wrong prepositions, articles, and particles	4	6	3	5	3	1	3	6	1	7	39
8		Using definite articles before genitives	0	3	0	0	1	0	0	0	0	0	4
9		Omitting functional morphemes (prepositions, articles, etc)	4	11	5	3	2	5	2	5	13	4	54
10	Semantic Errors	Using a wrong meaning of English homonyms	1	8	6	2	3	1	3	4	3	6	37
11		Using words of ambiguous meaning	0	0	3	0	0	3	0	0	0	0	6
12		Using terms that convey very different meaning	1	1	0	1	2	1	1	3	0	1	11
13		Using unfamiliar words in place of collocations	0	0	0	0	0	1	0	0	1	1	3
14		Using wrong reference and relative pronouns.	2	2	0	1	2	2	0	1	4	0	14
15		Adding an unnecessary word, preposition, or article before a word	10	3	1	5	4	4	8	9	6	1	51
16		Omitting necessary words or phrases	1	3	0	3	0	1	0	2	2	0	12
17		Corrupting the meaning of the whole sentence	1	1	1	0	0	0	0	0	0	0	3
Total			37	47	26	38	30	34	35	46	37	36	366
No. of sentences without errors			1	0	1	0	0	1	0	0	1	0	4