



## Analyzing Error Analysis in Google Translate: A Case Study of a Medical Text

Farah Abbas Abo Al-Timen [\*]

Lecturer at Department of Translation

College of Arts/ Al-Mustansiriyah University

[farah86@uomustansiriyah.edu.iq](mailto:farah86@uomustansiriyah.edu.iq)

### Abstract

Both sparing time and efforts urged a necessity to use machine translation (MT) systems such as Google Translate to translate texts all over the world. During the last decades, GT was the focus of attention for being the most used MT system for its speed and almost accurate product. This study emerged from the fact that English and Arabic translation via MT has many errors that need to be highlighted since both languages belong to different linguistic systems. In order to evaluate the quality of GT system, this study used error analysis with a view to identifying different error types in GT. The error analysis focused on studying the mismatching syntactic, morphological, lexical and semantic components in the source text and the target text. The case study for this research is an English medical text that is chosen to evaluate the accuracy and usefulness of GT in translating the English medical terms and statements. The results revealed that the most frequently occurring errors were semantic errors (30%) and syntactic errors (30%), morphological errors ranked second (25%) while lexical errors ranked last (15%).

**Keywords:** Arabic, English, Error Analysis, Google Translate, Machine Translation.

Farah Abbas Abo Al-Timen is a lecturer at the department of translation/ College of Arts [\*]  
– Al-Mustansiriyah University.

## تحليل الخطأ في ترجمة جوجل: دراسة حالة لنص طبي

م. فرح عباس أبو التمن

قسم الترجمة / كلية الآداب - الجامعة المستنصرية

### المستخلص

أدى توفير الوقت والجهود إلى ضرورة استخدام أنظمة الترجمة الآلية مثل ترجمة جوجل Google Translate لترجمة النصوص في جميع أنحاء العالم. كانت ترجمة جوجل خلال العقود الماضية محور الإهتمام لكونها أكثر أنظمة الترجمة الآلية استخدامًا وذلك لسرعتها والنتائج شبه الدقيقة التي أظهرتها. نشأت الحاجة لهذه الدراسة من حقيقة أن ترجمة الإنجليزية والعربية الآلية من خلال تطبيق جوجل يوجد فيها العديد من الأخطاء التي يجب تسليط الضوء عليها لأن كلتا اللغتين تنتمي إلى أنظمة لغوية مختلفة. من أجل تقييم جودة ترجمة جوجل، استخدم هذا البحث تحليل الخطأ بهدف تحديد أنواع الأخطاء المختلفة في ترجمة جوجل. ركز هذا التحليل على دراسة عدم تطابق المكونات النحوية والصرفية و المعجمية و الدلالية التي وجدت في النص المصدر و الهدف. دراسة الحالة لهذا البحث هي نص طبي إنجليزي تم اختياره لتقييم دقة وفائدة ترجمة جوجل في ترجمة العبارات و المصطلحات الطبية الإنجليزية. أظهرت النتائج أن الأخطاء الأكثر تكرارًا كانت أخطاء دلالية و نحوية بنسبة (٣٠٪) لكل خطأ بينما حلت الأخطاء المورفولوجية على المرتبة الثانية بنسبة (٢٥٪) أما الأخطاء المعجمية فقد حلت في المرتبة الأخيرة بنسبة (١٥٪). الكلمات المفتاحية: تحليل الخطأ، الترجمة الآلية، ترجمة جوجل، اللغة العربية، اللغة الإنجليزية.

### 1 Introduction

GT is one of the online MT software tools that serve "a multi-lingual MT translation service" (Li, Graesser, & Cai, 2014: 190). Great work has been devoted to the study of GT, the most well-known MT system in recent years (cf. Aziz, Sousa, & Specia, 2012; Karami, 2014 & Jabak 2019). Likewise an extensive research has been dedicated to evaluate MT and some experts have been concerned with using error analysis in their assessment of the machine systems, especially the system of Google (cf. Koponen, 2010; Stymne, 2011 & Popovic 2018). In order to identify the areas of errors in a particular language, a systematically contrastive analysis should be occurred between a language and its counterpart being translated by GT. Error Analysis is "the

alternative to contrastive analysis" of languages pairs (Rustipa, 2011: 18). Rustipa (ibid: 18) maintains that:

Error analysis is the study of kind and quantity of error that occurs, particularly in the fields of applied linguistics. These errors can be divided into three sub-categories: overgeneralization, incomplete rule application, and the hypothesizing of false concepts.

Though scientific language is more direct and free from alternative connotations, which can be found in a literary language, yet translating a scientific text can have some translation problems and hence producing some errors. Strevens in a survey article (1977: 153-4) maintains that:

Scientific discourse uses a lot of words, roots and affixes of Greek and Latin origin, and uses, or has access to, symbols, numbers and names of chemicals which are largely international in character.

Furthermore the style of the scientific language cannot be seen through the basic elements of the language only, but rather through "the statistical properties of the mixture in which they occur, and the intention, the purpose, behind their selection and use" (ibid: 153).

Accordingly, the translation of scientific texts is more than just dealing with terminology or just substituting a word with its equivalent in another language. Byrne (2006: 1) sheds light on this point by saying that:

The main concern for scientific translation is not only to make sure that information is conveyed accurately but they are also responsible for ensuring that the information is presented in the correct form, that it is complete and that the information can be used correctly and effectively.

Regardless of the enhancement that has been seen in GT system in recent years, the level of its accuracy is still rather debatable. The most important thing that should be focused in rendering scientific texts is to take into account the accuracy of GT in transferring the source language terms in one hand, and how to handle the terms in the target language on the other hand.

Therefore, this study can bring about some great improvements on how to translate properly by using a procedural error analysis.

## 2 Error Analysis: Literature Review

In order to evaluate MT, error analysis can be of a great help for the improvement of the translation system to measure the performance of the GT output since the word error means the imperfect use of words or grammatical items (Richards & Schmidt, 2002: 184). Errors are seen by many educators as an integral part of language teaching and learning (Aqel & Mohammed, 2017: 127). Corder (1974: 125) stated that "the study of errors is part of the investigation of the process of language learning." Crystal (2003: 165) also attached the study of errors to language learning by mentioning that error analysis is considered as a method for identifying the imperfect forms produced by a foreign language learner, or in this case by a machine, under the principles and procedures provided by linguistics.

Perspectives in translation studies recognized error analysis as a fundamental tool to the process of translation, since it has the ability to develop the outcome of translation (Van der Wees et al., 2015). Approaches for error classification appeared by using the application of GT to compare the translation product with a reference translation. For comparing MT outcomes for different languages, error has been classified into different types. Farr'us et al. (2010) defined a simple error scheme comprising five classes for comparing the Spanish-Catalan languages in both directions: (1) morphological errors, (2) lexical errors, (3) syntactic errors, (4) orthographic errors and (5) semantic errors. Federico et al. (2014) gave a set of basic error classes for analyzing MT product from English into different language pairs involving Arabic, Chinese and Russian: (1) morphological errors, (2) lexical choice (additions and omissions), (3) casing and punctuation and (4) word order errors. Similar basic scheme (without "casing and punctuation") was used in Castilho et al. (2017) to compare phrase-based MT outputs for a number of language pairs and set the following error types: (1) morphology, (2) word order, (3) omission, (4) addition and (5) mistranslation. As for Costa et al., they (2015) gave a comparison of five different error classes of English

to Portuguese systems: (1) orthography, (2) lexis, (3) grammar, (4) semantic and (5) discourse. While Kirchhoff et al. (2012) offered a more detailed error classes for comparing English-to-Spanish translations: (1) missing words, (2) extra words, (3) word order, (4) morphology, (5) word (sense/ term), (6) capitalization, (7) punctuation, (8) pragmatics and (9) diacritics.

Given all of the above mentioned classifications of error, the research analysis focused on studying the English and the Arabic texts through the main four errors of a language:

- Morphological errors (derivation - word formation, inflection, number and person).
- Lexical errors (addition and omission of words, conjunctions, etc.).
- Syntactic errors (word order, passive voice, noun phrase, tense, articles, subject-verb agreement, conjunction, relative clause, etc.).
- Semantic errors (semantic relations, collocation, word sense, etc.).

### 3 Method

This study is descriptive and qualitative and the method used for the analysis is error analysis. This method is used for analyzing the GT errors at the syntactic, morphological, lexical and semantic levels. At first the application (GT) was chosen to be the reference translation to compare its Arabic translation output with the source English text. Secondly, the translated text was analyzed through its sentences, phrases and words. Then, each error was examined and organized in tables and analyzed at the four levels.

### 4 Case Study

The case study for this research is an English medical text about Corona Virus<sup>1</sup> and its Arabic version translated by GT as they are listed below in table (1) and (2).

### Corona Virus

Coronaviruses are a group of related RNA viruses that cause diseases in mammals and birds. In humans and birds, they cause respiratory tract infections that can range from mild to lethal. Mild illnesses in humans include some cases of the common cold (which is also caused by other viruses, predominantly rhinoviruses), while more lethal varieties can cause SARS, MERS, and COVID-19. In cows and pigs they cause diarrhea, while in mice they cause hepatitis and encephalomyelitis. Generally, there were no vaccines or antiviral drugs available for use against coronaviruses until the COVID-19 pandemic which broke out in late 2019.

The name "corona virus" is derived from Latin corona, meaning "crown" or "wreath", itself a borrowing from Greek κορώνη korōnē, "garland, wreath". The name was coined by June Almeida and David Tyrrell who first observed and studied human coronaviruses.

Table (1)

The Original English Text

فيروس كورونا
فيروسات كورونا هي مجموعة من فيروسات الحمض النووي الريبي ذات الصلة والتي تسبب الأمراض في الثدييات والطيور. في البشر والطيور، تسبب التهابات في الجهاز التنفسي يمكن أن تتراوح من خفيفة إلى مميتة. تشمل الأمراض الخفيفة التي تصيب البشر بعض حالات نزلات البرد (التي تسببها أيضاً فيروسات أخرى، في الغالب فيروسات الأنف)، في حين أن الأنواع الأكثر فتكاً يمكن أن تسبب السارس، وفيروس كورونا، و COVID-19. في الأبقار والخنزير تسبب الإسهال، بينما تسبب في الفئران التهاب الكبد والتهاب الدماغ. بشكل عام، لم تكن هناك لقاحات أو عقاقير مضادة للفيروسات متاحة للاستخدام ضد فيروسات كورونا حتى جائحة COVID-19 التي اندلعت في أواخر عام ٢٠١٩. اشتق اسم "فيروس كورونا" من اللغة اللاتينية كورونا، وتعني "تاج" أو "إكليل"، وهي نفسها مستعارة من اللغة اليونانية κορώνη korōnē ، "إكليل، إكليل". صاغ الاسم جون ألميدا وديفيد تيريل أول من لاحظ ودرس فيروسات كورونا البشرية.

Table (2)

The Target Text Translated by GT

## 5 Discussing of Data Findings

The analysis showed that GT has many errors at the syntactic, morphological, lexical and semantic levels as it is clear from the tables and the analysis below:

No	Source Text	Syntactic Errors of the Target Text		
		Word order	Relative Clauses	Subject-Verb Agreement
1	that cause diseases		التي تسبب	
2	In humans and birds, they cause respiratory tract infections	في البشر والطيور، تسبب التهابات في الجهاز التنفسي		
3	In cows and pigs they cause diarrhea,	في الأبقار والخنازير تسبب الإسهال		
4	a group of related RNA viruses	مجموعة من فيروسات الحمض النووي الريبي ذات الصلة		
5	June Almeida and David Tyrrell who first observed		جون ألميدا وديفيد تيريل أول من لاحظ	
6	June Almeida and David Tyrrell who first observed			جون ألميدا وديفيد تيريل أول من لاحظ

Table (3)

### Syntactic Errors Analysis

No.	Source Text	Morphological Errors of the Target Text		
		Number	Personal Pronoun	Derivation (Word Formation)
7	which is also caused by other viruses, predominantly rhinoviruses		التي تسببها أيضا فيروسات أخرى، في الغالب فيروسات الأنف	
8	MERS			فيروس كورونا
9	COVID-19			COVID-19
10	a borrowing from Greek κορόνη korōnē	مستعارة من اللغة اليونانية κορόνη korōnē		
11	itself a borrowing from Greek		وهي نفسها مستعارة من اللغة اليونانية	

Table (4)

### Morphological Errors Analysis

No	Source Text	Lexical Errors of the Target Text	
		Addition	Omission
12	Coronaviruses are a group of related RNA viruses that cause diseases in mammals and birds.	فيروسات كورونا هي مجموعة من فيروسات الحمض النووي الريبي ذات الصلة والتي تسبب الأمراض في الثدييات والطيور.	
13	respiratory tract infections that can range from mild to lethal		التهابات في الجهاز التنفسي يمكن أن تتراوح من خفيفة إلى مميتة
14	MERS		فيروس كورونا

**Table (5)**

**Lexical Errors Analysis**

No.	Source Text	Semantic Errors of the Target Text		
		Semantic Relations (synonymy, polysemy, homonymy, etc.)	Collocation	Word Sense
15	infections that can range from mild to lethal.		التهابات يمكن أن تتراوح من خفيفة إلى مميتة.	
16	mild illnesses		أمراض خفيفة	
17	broke out		اندلعت	
18	a borrowing from Latin corona			مستعارة من اللغة اللاتينية كورونا
19	garland	اكليل		
20	human coronaviruses			فيروسات كورونا البشرية

**Table (6)**

**Semantic Error Analysis**

In table (3), GT shows syntactic errors at the levels of word-order, relative clause and subject-verb agreement as in the following examples: in [1], the relative clause *a group of related RNA viruses that cause diseases* is translated into مجموعة من فيروسات الحمض النووي الريبي ذات الصلة والتي تسبب الأمراض in which the word *group* is preceded by an indefinite article *a*, therefor, the relative pronoun should be omitted from the target text and the whole relative clause should be translated into مجموعة من فيروسات... (تسبب) أمراضا. In [2], [3] and [4], there are examples of word order error: in [2], *In humans and birds, they cause respiratory tract infections* is rendered into في البشر والطيور، تسبب whereas it should be translated into تسبب إنتهابات في الجهاز التنفسي theهابات تسبب إنتهابات في الجهاز التنفسي في البشر و الطيور in [3], *In cows and pigs they cause diarrhea* became في تسبب الإسهال في whereas it should be translated into تسبب الإسهال في الأبقار والخنازير and in [4], *a group of related RNA viruses* the sentence became مجموعة من فيروسات الحمض النووي الريبي ذات الصلة whereas it should be translated into فيروسات ذات صلة بفيروسات الحمض النووي الريبي the sentence became فيروسات ذات صلة بفيروسات الحمض النووي الريبي and the definite article *أل* should be omitted from الصلة. In [5], there is a relative clause error in *June Almeida and David Tyrrell who first observed* that is rendered into جون ألميدا وديفيد تيريل أول من لاحظ in which the relative pronoun *who* that refers to *June* and *David* is omitted and it should be translated into اللذان. In [6], with the

same previous example mentioned in [5], there is also a subject-verb agreement error in which there is no agreement between the two subjects *June* and *David* and the verb *observed* and the whole sentence should be translated into جون و ديفيد اللذان (كانا) أول من لاحظ

In table (4) GT shows morphological errors that are seen through number, personal pronoun and word formation. For example in [7], *which is also caused by other viruses, predominantly rhinoviruses* is rendered into التي تسببها التي تسببها أيضا فيروسات أخرى، في الغالب فيروسات الأنف while there should be an addition of the personal pronoun هي before the غالب : في الغالب (هي) في : في الغالب... In [8], the rendering of the derived abbreviation *MERS* shows a morphological error in which GT mistranslates the initial letters which mean: *Middle East respiratory syndrome-related coronavirus*, therefore, they should be rendered into فيروس كورونا المرتبط بمتلازمة الشرق الأوسط التنفسية. Also the abbreviation *COVID-19* in [9] is rendered into *COVID-19* while it should be translated into كوفيد-19 or كورونا فيروس. In [10], there is a number error in rendering the word *Greek* that should be translated into الكلمتين اليونانيتين instead of اللغة اليونانية since there are two words to be described after the word *Greek*: *κορώνη* and *korōnē*. In [11], there is a personal pronoun error in translating *itself a borrowing from Greek* in which the pronoun *itself* is rendered into هي نفسها while it should be translated into نفسها الكلمة because the pronoun refers to *corona*.

The lexical error is also detected through the addition and omission examples as it is shown in table (5). For example, in [12], there is an addition error in rendering *Coronaviruses are a group of related RNA viruses that cause diseases in mammals and birds* that is translated into فيروسات كورونا هي مجموعة من فيروسات الحمض النووي الريبي ذات الصلة والتي تسبب الأمراض في الثدييات والطيور in which the conjunction و cannot be preceded before a relative pronoun. In [13], there is an omission between تتراوح and من in rendering *that can range from mild to lethal* into , تتراوح من نخيفة إلى مميتة , in which it should be translated into تتراوح شدتها/ شدة اعراضها من خفيفة إلى مميتة. In [14], there is also an omission in rendering the abbreviation *MERS* in which it is rendered

المرتبط بمتلازمة while the whole correct translation should be المرتبط بمتلازمة الشرق الأوسط التنفسية.

Lastly, it can be seen from table (6) that there are semantic errors at the levels of collocation, word sense and synonymy as in the following examples: there is a collocation error represented in the translation of the word *mild* in which in the English text it is collocated with the word *infections* in [15] and the word *illness* in [16] and in it is rendered in the Arabic text in both cases into خفيفة while it should be translated into طفيفة in order to be collocated with أمراض in [15] and التهابات in [16]: أمراض طفيفة، التهابات طفيفة. Another example of collocation error is found in [17] in which *pandemic* is collocated with *broke out* and in Arabic it is rendered into جائحة اندلعت because كوفيد-19 انتشرت or ظهرت while it should be translated into نار or حرب but cannot be collocated with مرض. Word sense error can be seen in [18] in the word *Latin* in: *The name corona virus is derived from Latin corona* that is rendered into اللغة اللاتينية while it should be translated into الكلمة اللاتينية because it is a description for a word *corona* not a whole language. In [19], there is a semantic relation error of synonymy represented in the translation of the word *garland* in which this word and the word *wreath* are both rendered into إكليل while *garland* should be translated into تاج. Another word sense error is found in [20] in the translation of the word *human* in: *human coronaviruses* that is rendered into فيروسات كورونا البشرية in which *corona* is given a human quality بشرية whereas it should be rendered with a relative clause: (التي تصيب البشر).

The final analysis of the frequently occurring errors is shown in table (7) below:

No.	Analysis Errors	Frequency Data	Percentage
1	Syntactic Errors	6	30 %
2	Morphological Errors	5	25 %
3	Lexical Errors	3	15 %
4	Semantic Errors	6	30 %
	Total	20	100 %

**Table (7)**

The Frequency Occurring of Errors

To compare the present study results with the previous five ones mentioned in the literature review, it is evident that GT shows mismatches in the different linguistic levels: semantic, morphological, lexical and syntactic in all the studies. It could be stated that some levels (the syntactic and the semantic levels) show more errors than other error types as it is listed in table (8) below.

Study Name	Error Types Results
Farr'us et al (2010)	It was observed that the semantic errors have the highest frequencies than other error types (37%). Syntactic errors have also a high frequency (33%). Morphological and lexical errors have the lowest frequency (15%) for each type.
Kirchhoff et al. (2012)	The results indicated that the syntactic errors were the most frequent error type (35%), followed by semantic errors (30%). Morphological errors (20%) and lexical errors (15%) were the less frequent error types.
Federico et al. (2014)	The obtained results showed that the highest frequencies of errors were observed on the semantic and syntactic levels (36%) for each type. Fewer lexical errors (18%) and morphological errors (10%) were found.
Costa et al. (2015)	It was shown that semantic (34%) and syntactic (32%) errors have the most frequencies. Lexical errors (24%) were the second most frequent error type. Low number of morphological errors (10%) were found.
Castilho et al. (2017)	<p>he largest correlation is observed for lexical errors and missing words. Additional very interesting finding is that the human perception of quality does not necessarily depend on frequency of the given error type – a sentence with a low overall score can easily contain less missing words and/or lexical errors than an- other sentence with higher score</p> <p>he largest correlation is observed for lexical errors and missing words. Additional very interesting finding is that the human perception of quality does not necessarily depend on frequency of the given error type – a sentence with a low overall score can easily contain less missing words and/or lexical errors than an- other sentence with higher score</p> <p>The total count of errors showed that semantic (33 %) and syntactic (31%) errors were the highest while fewer morphological (20%) and lexical (16%) errors were found.</p> <p>errors of omission, addition, and mistranslation</p>
The Present Study (2021)	The results showed that semantic and syntactic errors have high ranks between other error types (30%) for each type. Morphological errors have ranked third (25%) while lexical errors ranked last (15%).

### Table (8)

Comparing the Present Study Results with other Studies

These results with previous data findings reveal the reliability of using error analysis as a vital tool of improving MT output and also suggest that GT needs to have a human intervention to apply some modifications on its output in order to provide an adequate result as that obtained by a human translation.

### 6 Conclusions

The main focus of the research is to analyze the quality of GT and introduce sufficient insight into the classifications of errors and their frequencies to achieve an automatic assessment with the intention of improving GT system. From the analysis of the English medical text and its Arabic translation via GT, the data findings show that the most frequently occurring errors are semantic and syntactic, followed by the morphological errors and the least frequently occurring errors are lexical errors. These frequencies indicate that GT doesn't live up to an adequate translation and it can be considered as a draft translation that needs to have some adjustments in order to get an acceptable result. Accordingly, the revealed errors detected by this research can be used for teaching and learning how to make the translation of Google useful for its users. However, it cannot be denied that the translation of GT is a worthy aid to human translation due its speed in performing translation tasks worldwide.

Furthermore, machine translation field is still an unknown arena that needs further studies and requires more efforts to investigate its system. This research may help the experts and developers of GT to be aware of the weaknesses points found in the system and observe them carefully in order to enhance future GT versions and avoid any former errors.

Additionally, machine translation is an unknown field of study in Iran and needs a lot of efforts to be investigated.

This study, beside other research done in Iran, may help experts to write better computer programs.



Machine translations, as aids to human translation besides the vast development of technology in using computers, have brought machine translation evaluation into consideration. The quality investigation of Google Translate as a machine translation system and the analysis of its weaknesses were to light a number of ideas to improve future made softwares and help users to adjust their expectations and have better understanding. The findings are of direct practical relevance

Machine translations, as aids to human translation besides the vast development of technology in using computers, have brought machine translation evaluation into consideration. The quality investigation of Google Translate as a machine translation system and the analysis of its weaknesses were to light a number of ideas to improve future made softwares and help users to adjust their expectations and have better understanding. The findings are of direct practical relevance

Machine translations, as aids to human translation besides the vast development of technology in using computers, have brought machine translation evaluation into consideration. The quality investigation of Google Translate as a machine translation system and the analysis of its weaknesses were to light a number of ideas to improve future made softwares and help users to adjust their expectations and have better understanding. The findings are of direct practical relevance

Machine translations, as aids to human translation besides the vast development of technology in using computers, have brought machine translation evaluation into consideration. The quality investigation of Google Translate as a machine translation system and the analysis of its weaknesses were to light a number of ideas to improve future made softwares and help users to adjust their expectations and have better understanding. The findings are of direct practical relevance

## References

Aqel, K. A. J. D., & Mohammed, F. (2017). *A Longitudinal Analysis Study of Writing Errors Made by EFL*. British Journal of Education Vol.5, issue 5, pp. 127-145.

Aziz, W., Sousa, S. C. M., & Specia, L. (2012). PET: A tool for post-editing and assessing machine translation. In *The Eighth International Conference on Language Resources and Evaluation*, Istanbul, Turkey.

Byrne, J. (2006). *Technical translation: Usability strategies for translating technical documentation*. Dordrecht: Springer.

Castilho S, Moorkens J, Gaspari F, Sennrich R, Sosoni V, Georgakopoulou P, Lohar P, Way A, Barone AVM, Gialama M (2017). A Comparative Quality Evaluation of PBSMT and NMT using Professional Translators. In: *Proceedings of MT Summit XVI*. Vol. 108, issue 1, pp. 116–131.

Corder, S.P. (1974). Error Analysis, In Allen, J. L. P. and Corder, S. P. (1974). *Techniques in Applied Linguistics*. Oxford: Oxford University Press.

Costa A, Ling W, Lu'is T, Correia R, Coheur L (2015). A linguistically motivated taxonomy for Machine Translation error analysis. In *Machine Translation Journal*. Vol. 29, issue 2, pp:127–161.

Crystal, D. (2003). *A Dictionary of Linguistics and Phonetics*. 5<sup>th</sup> ed. Blackwell Publishing Ltd.

Farr'us M, Costa-Juss'a MR, Mari'no JB, Fonollosa JAR (2010). Linguistic-based Evaluation Criteria to Identify Statistical Machine Translation Errors. In: *Proceedings of the 14th Annual Conference of the European Association for Machine Translation (EAMT 2010)*, Saint-Raphal, France, pp. 167–173.

Federico M., Negri M., Bentivogli L & Turchi, M. (2014). Assessing the Impact of Translation Errors on Machine Translation Quality with Mixed-effects Models. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, Doha, Qatar, pp. 1643–1653.

Jabak, O. (2019). Assessment of Arabic-English translation produced by Google Translate. *International Journal of Linguistics, Literature and Translation (IJLLT)*, Vol. 2, Issue 4, pp. 238-247.

Karami, O. (2014). The Brief View on Google Translate Machine. *Paper presented at the meeting of the 2014 Seminar in Artificial Intelligence on Natural Language, German*.

Retrieved from: <https://www.semanticscholar.org/paper/The-brief-view-on-Google-Translate-Machine-Karami/c6f95d543c0b34c4026b9e6cf64decd94b793823>

Kirchhoff K, Capurro D, Turner A (2012). Evaluating user preferences in machine translation using conjoint analysis. In: *EAMT 2012: Proceedings of the 16th Annual Conference of the European Association for Machine Translation, Italy*, pp. 119–126.

Koponen, M. (2010). Assessing machine translation quality with error analysis. In *MikaEL: Electronic Proceedings of the KäTu Symposium on Translation and Interpreting Studies*, pp. 1-12. Retrieved from: [https://sktl-fi.directo.fi/@Bin/40701/Koponen\\_MikaEL2010.pdf](https://sktl-fi.directo.fi/@Bin/40701/Koponen_MikaEL2010.pdf)

Li, H., Graesser, A. C., & Cai, Z. (2014). Comparison of Google translation with human translation. In *27th International Florida Artificial Intelligence Research Society Conference*, pp. 190–195. Association for the Advancement of Artificial Intelligence.

Popovic, M. (2018). Error Classification and Analysis for Machine Translation Quality Assessment. J. Moorkens, et al. (eds.), In *Translation Quality Assessment*, pp. 129-158. Springer International Publishing.

Richards, J. C. & Schmidt, R. (2002). *Dictionary of language teaching and applied linguistics*. 3<sup>rd</sup>, ed. London: Longman.

Rustipa, K. (2011). Contrastive Analysis, Error Analysis, Interlanguage and the Implication to Language Teaching. Stikuban University (Unisbank) Semarang. *Ragam Jurnal Pengembangan Humaniora*, Vol. 11, pp. 16-22. Retrieved from: <https://www.semanticscholar.org/paper/Contrastive-Analysis-%2C-Error-Analysis-%2C-and-the-to-Rustipa/255ebdf90f8353aac8347ec8fcbe7e0f66fa92f2>

Stevens, P. (1977). *Special Purpose Language learning: a perspective*. Language Teaching & Linguistics Abstracts Vol. 10 (3), pp. 145-163.

Stymne, S. (2011). Blast: A tool for error analysis of machine translation output. In *Proceedings of ACL*, pp. 56-61. Portland, Oregon, USA.

Van der wees, M., Bissaza, A. & Monz, C. (2015). Five Shades of Noise: Analyzing Machine Translation Errors in User-Generated Text. In *Proceedings of the Workshop on Noisy User-generated Text*, pp. 28–37. Retrieved from [https://www.researchgate.net/publication/301449172\\_Five\\_Shades\\_of\\_Noise\\_Analyzing\\_Machine\\_Translation\\_Errors\\_in\\_User-Generated\\_Text](https://www.researchgate.net/publication/301449172_Five_Shades_of_Noise_Analyzing_Machine_Translation_Errors_in_User-Generated_Text)

Widdowson, H.G. (1974). Literary and Scientific Uses of English. In *English Language Teaching Journal*. Vol. 3 (28), pp. 282-292.

It is shown that lexical and semantic errors have

most impact on sentence level ranking. Furthermore, highly ranked sentences clearly exhibit low number of grammatical errors, however the relation between grammatical errors and poorly ranked segments remained unclear. Apart from this, high inter-annotator agreement between two annotators is reported, which contradicts the results from former studies. The most probable factor is removing words with position disagreement from calculation which increased the agreement between the error types



is shown that lexical and semantic errors have most impact on sentence level ranking. Furthermore, highly ranked sentences clearly exhibit low number of grammatical errors, however the relation between grammatical errors and poorly ranked segments remained unclear. Apart from this, high inter-annotator agreement between two annotators is reported, which contradicts the results from former studies. The most probable factor is removing words with position disagreement from calculation which increased the agreement between the errrgest correlation is observed for lexical errors and missing words. Additional very interesting finding is that the human perception of quality does not necessarily depend on frequency of the given error type – a sentence with a low overall score can easily contain less missing words and/or lexical errors than another sentence with