

A comparative study on google translate: An error analysis of Turkish-to English translations in terms of the text typology of Katherina Reiss

Betül KOÇER GÜLDAL¹

Korkut Uluç İŞİSAĞ²

APA: Koçer Güldal, B.; İşisağ, K. U. (2019). A comparative study on google translate: An error analysis of Turkish-to English translations in terms of the text typology of Katherina Reiss. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*, (Ö5), 367-376. DOI: 10.29000/rumelide.606217.

Abstract

Machine Translation (MT) has become one of the important topics of public interest especially with the advent of technology and the blooming need for translation. *Google Translate*, as an MT, provides quick translations; however, the quality of the texts often remains unsatisfactory. This study aims to analyze the translation errors of *Google Translate* outputs conducted from Turkish into English. The errors are classified into four major categories: *Lexical Errors*, *Morphological Errors*, *Syntactic Errors*, *Semantic and Pragmatic Errors*, which include subcategories. In parallel with the aim of the study, a text from each of the three text types put forward by Katherina Reiss (1971), was chosen to be translated by Google Translate and to be analyzed. These text types are *Informative Texts*, *Expressive Texts*, and *Operative Texts*. In the study, firstly it is aimed to explore which of the main text types has more translation errors, secondly, whether the translation error types vary by the main text types or not. In order to deal with this, both quantitative and qualitative analyses are utilized in the study. The data analysis revealed that the main text type that has more translation errors is the translation of *operative text* and *expressive text*, respectively. It is also observed that the error pattern between the text types was different. The informative text mainly includes *lexical errors*, whereas operative and expressive mainly include *semantic* and *pragmatic errors*. Summing up the results, it can be concluded that although *Google Translate* provides much quicker translations among a large number of languages, there is still a need for human assistance.

Keywords: Google translate, machine translation, translation errors, text typology, translation accuracy.

Google çeviri üzerine karşılaştırmalı bir çalışma: Katherina Reiss'in metin tipolojisi açısından Türkçe-İngilizce çevirilerin hata analizi

Öz

Makine Çevirisi (MÇ), özellikle teknolojinin ortaya çıkması ve çeviri ihtiyacının artması ile birlikte halkın ilgilendiği önemli konulardan biri haline gelmiştir. *Google Translate*, MÇ olarak hızlı çeviri sağlar; ancak, metinlerin kalitesi çoğu zaman yetersiz kalmaktadır. Bu çalışma, Türkçeden İngilizceye yapılan Google Translate çıktılarının çeviri hatalarını incelemeyi amaçlamaktadır. Hatalar, alt kategorileri olan dört ana kategoriye ayrılır: *Sözlüksel Hatalar*, *Morfolojik Hatalar*, *Sözdizimsel Hatalar*, *Anlamsal ve Pragmatik Hatalar*. Çalışmanın amacı doğrultusunda, Katharina Reiss (1971)

1 Arş. Gör., Ankara Hacı Bayram Veli Üniversitesi, Edebiyat Fakültesi, Mütercim Tercümanlık Bölümü, İngilizce Mütercim Tercümanlık ABD (Ankara, Türkiye), betulkocer@gmail.com, ORCID ID: 0000-0002-9306-5561 [Makale kayıt tarihi: 20.07.2019-kabul tarihi: 19.08.2019; DOI: 10.29000/rumelide.606217]

2 Dr. Öğr. Üyesi, Ankara Hacı Bayram Veli Üniversitesi, Edebiyat Fakültesi, Mütercim Tercümanlık Bölümü, İngilizce Mütercim Tercümanlık ABD (Ankara, Türkiye), kisisag@gmail.com, ORCID ID: 0000-0002-3569-4669.

tarafından öne sürülen her üç metin türünden birer metin Google Translate tarafından çevrilmek ve incelenmek üzere seçilmiştir. Bu metin türleri Bilgilendirici Metinler, Anlatımcı Metinler ve İşlevsel Metinlerdir. Çalışmada, ilk olarak, ana metin türlerinden hangisinde daha fazla çeviri hatası bulunduğu, ikinci olarak da, çeviri hatası türlerinin ana metin türlerine göre değişip değişmediğinin araştırılması amaçlanmıştır. Bu amaçla, çalışmada hem nicel hem de nitel analizler kullanılmıştır. Veri analizi, çeviri hatası daha fazla olan ana metin türünün sırasıyla işlevsel metin ve anlatımcı metin çevirisi olduğunu ortaya koymuştur. Metin tipleri arasındaki hata türlerinin farklı olduğu da gözlenmiştir. Bilgilendirici metin temel olarak sözcüksel hatalar içerirken, işlevsel ve anlatımcı metinler temel olarak anlamsal ve pragmatik hatalar içermektedir. Sonuçları özetlerken, Google Translate'in çok sayıda dilde daha hızlı çeviriler sağlamasına rağmen, hala insan yardımına ihtiyaç duyulduğu sonucuna varılabilir.

Anahtar kelimeler: Google Translate, makine çevirisi, çeviri hataları, metin tipolojisi, çeviri doğruluğu.

1. Introduction

Machine Translation (MT) has become one of the important topics of public interest especially with the advent of technology and the blooming need for translation. The initial goal of machine translation was to build an automatic high-quality machine translation that did not require any human assistance. However, Bar-Hillel reported that building a fully automatic translation system was unrealistic (as cited in Quah, 2006, p. 7) and years later the translation quality of MT such as *Google Translate*, still remained controversial, although the term originally referred only to “automatic systems with no human involvement” (Sager, 1994, p. 326).

It has been constantly pointed out by scholars that any process of translation has to result in a translation of good quality. With the aim to evaluate the quality of a translation, many scholars (Pym, 1993; House, 1997; Lauscher, 2000; Mossop, 2014) suggest error analysis between source text (ST) and target text (TT). In recent years, research on machine translation evaluation has become very popular and some experts have been interested in using error analysis to assess machine translations (Eftekhari & Nouraei, 2013; Koponen, 2010; Stymne, 2011).

Error analysis of *Google Translate* has been conducted by many researchers (Jamillah, 2009; Aiken & Balan, 2011; Balk et al., 2012; Ghasemi & Hashemian, 2016), especially the research of Aiken and Balan (2011) offered an insight into studies on *Google Translate* translations, they assessed the translation quality of *Google Translate* considering 50 different languages, consequently they reached the conclusion that *Google Translate* translates a European language into another European language much better than other languages. However, there have been no academic, comprehensive translation quality assessments of *Google Translate* in Turkish-English language pairs.

The aim of the present study is to evaluate the data according to different text types and to contribute to the discussion about translation quality of MT by providing another piece of analysis of translation errors in Turkish-to-English translation of *Google Translate*, which is an MT providing service to translate different written texts from one language to another and it provides translating over 90 languages. In the study, firstly it is aimed to explore which of the main text types has more translation errors, secondly, whether the translation error types vary by the main text types or not. In order to deal with this, both quantitative and qualitative analyses are utilized in the study. *Google Translate* search

different documentaries to find the best appropriate translation pattern between translated texts by a human. Thus, the quality of *Google Translate* depends on the number of human translated texts searched by Google Translate (Karami, 2014). Karami also discussed different models used in Google Translate. He focused on two major engines used by Google Translate: rule-based and statistical based. The second is used in Turkey as free, which is the tool of this study.

2. Methodology

2.1. Data collection

Both quantitative and qualitative analyses are utilized in the study. As the focus of the study is on translation errors, a comparative and descriptive error analysis between the ST and the TT is conducted as the main method of data analysis. To assure the reliability of the results obtained in the data analysis, *the inter-rater reliability* is provided with two professional translators and academicians. The agreement between the experts and the researcher is reached. In parallel with the aim of the study, an example text from each of the three text types put forward by Katherina Reiss (1971), was chosen to be translated by Google Translate and to be analyzed. These text types are *Informative Texts*, *Expressive Texts*, and *Operative Texts*. The source texts have been chosen from the field of *translation and language* so that their terminologies could be analyzed consistently. For Informative Texts, an online course excerpt about translation which is compiled from Sakine Eruz's translation class, for Expressive Texts, the poem of Bedri Rahmi Eyüboğlu titled "3 Dil" [*3 Languages*], and for Operative Texts, the slogan of ÇEVBİR, which is "Biz Çevirmezsek, Dünya Dönmez" have been chosen to analyze.

2.2. Data analysis

The source texts analyzed in the study have been classified in accordance with the text typology presented by Katharina Reiss, which is developed based on Karl Bühler's Organon Model (1971). The data has been analyzed in terms of *Lexical*, *Morphological*, *Syntactic*, *Semantic*, and *Pragmatic Errors*. All the translation processes of the source texts have been conducted on the date May 28, 2018 through *Google Translate in Turkey* and their screenshots have been saved.

2.3. Error types

Keshavarz (1999) linguistically divided errors into five major groups which are Lexical Errors, Morphological Errors, Syntactic Errors, Semantic errors, and Pragmatic Errors.

In the present study, comparative error analysis has been conducted in terms of four main translation error types which are Lexical Errors, Morphological Errors, Syntactic Errors, and Semantic and Pragmatic Errors.

Lexical errors refer to mistakes at the word level (Hernández, 2011, p. 264). On the basis of the definition, when the TT reflected errors in wrong word choice, literal translation, omission, addition, errors in word formation, the errors were categorized under lexical errors in the present study.

Morphology is the study of morphemes, which are the smallest significant units of grammar (Todd, 1987). Accordingly, morphological errors refer to mistakes at the function of morphemes. In the present study; affixation related errors, compound-related errors, failure to use the marker (-er), conversion related errors, using of plural –s morpheme were categorized under morphological errors.

The syntax is the set of rules, principles, and processes that govern the structure of sentences in a given language, usually including word order (Chomsky, 1957, p. 11). Accordingly, syntactic errors refer to errors in the arrangement of words or phrases of a sentence. Errors in using auxiliaries, errors in using SVO pattern, errors in using articles, errors in using prepositions, errors in using the correct form of a tense, and errors in using conjunctions were categorized under syntactic errors.

Semantics is the study of meaning. James (1998) classifies semantic errors into two types: Confusion of sense relations and collocation errors. Confusion of sense errors encompasses four types of errors: (1) using a superonym for a hyponym, (2) using a hyponym for a superonym, (3) using inappropriate co-hyponyms, (4) using a wrong near synonym. Collocation is a word or phrase that is frequently used with another word or phrase. Accordingly, the ambiguity in translation related to the confusion of sense relations and collocation errors; and pragmatic errors that undermine the meaning were categorized under semantic and pragmatic errors.

2.4. Statistical techniques

A statistical test has been conducted to determine whether the error means are statistically different from each other or not.

The test statistics W is defined as follows (Gu et al., 2008).

$$W = \frac{X_1 - X_2/d}{\sqrt{X_1 + X_2/(d^2)}}$$

Since the p-value is 0.0000 and less than the significance level 0.01, the null hypothesis is rejected.

Data provides sufficient evidence to conclude that there is a significant difference between the means of translation errors in text types.

3. Results and discussion

In analysis of translation errors of Google Translate in the Turkish-to-English translation, for informative text type, 904 lexical items were examined; translation errors were identified and classified according to their error types. The data analysis revealed that there were 38 incorrect instances of translation, which is equal to 4.2% of all lexical items revised. For expressive text type, 163 words were examined; there were 26 incorrect instances of translation, which is equal to 15.95% of all lexical items revised. For operative text type, 4 words were examined; there were 2 incorrect instances of translation, which is equal to 50% of all lexical items revised. Table 1 below presents the summary of error types and their quantities.

Table 1. Summary of error types and their quantities.

	Informative Text	Expressive text	Operative Text
Lexical Errors	16 42.1%	4 15.4%	1 50%
Morphological Errors	3 7.9%	-	-
Syntactic Errors	14 36.9%	2 7.8%	-
Semantic Errors	3 7.9%	10 38.4%	-
Pragmatic Errors	2 5.2%	10 38.4%	1 50%
Total	38	26	2

As it is evident from Table, among the 66 total instances of translation errors in the source texts, from total 38 instances of informative text, 16 were defined as Lexical Errors, 3 as Morphological Errors, 14 as Syntactic Errors, 3 as Semantic Errors, and 2 as Pragmatic Errors. From total 26 instances of expressive text, 4 were defined as Lexical Errors, no Morphological Errors, 2 as Syntactic Errors, 10 as Semantic Errors, and 10 as Pragmatic Errors. From total 2 instances of operative text, 1 was Lexical Error and the other 1 was Pragmatic Error.

After the quantities of the errors in translation are determined and categorized, each type of them is defined and discussed separately below.

3.1. Analysis of informative text

Among the 38 total instances of translation errors in the informative text, 16 instances were defined as Lexical Errors, 3 as Morphological Errors, 14 as Syntactic Errors, 3 as Semantic Errors, and 2 as Pragmatic Errors. One example that has been chosen from each error types below is illustrative.

Example 1: A Lexical Error in the Informative Text

<i>ST</i>	<i>Çoğul dizge kuramı</i>
<i>MT</i>	<i>Theory of Multiple Lines</i>
<i>Revision</i>	<i>Polysystem Theory</i>

In Example 1, the statement of *Çoğul dizge kuramı* [*Polysystem Theory*] was translated literally as *Theory of Multiple Lines* which cannot convey the intended meaning of the source sentence. This error was classified under lexical errors for being word-for-word translation.

Example 2: A Morphological Error in the Informative Text

<i>ST</i>	<i>Çağdaş Çeviribilimsel yaklaşımlar yetmişli yıllarda çevirinin özgül konumuna uygun şekillenmeye başlamıştır</i>
<i>MT</i>	<i>Contemporary Theories of translational science have begun to shape in the mid-seventy for the specific position of translation.</i>
<i>Revision</i>	<i>Mid-seventies</i>

In Example 2, the phrase *yetmişli yıllar* [seventies] are in plural form in the ST, however, the –s plural suffix was omitted in the MT and translated as a singular word. This error was classified under morphological errors for presenting an error in the plurality of the word.

Example 3: A Syntactic Error in the Informative Text

ST	<i>Kaldı ki, öncelikle yirminci yüzyılın ilk yarısından itibaren çeviri metin türleri çeşitlenmiş ve çeviri olgusu teknolojik gelişmelerle de erek kültürde yerleşmiştir</i>
MT	<i>Moreover, since the first half of the twentieth century, the types of translation texts have been diversified and the translation phenomenon has become localized in technological developments</i>
Revision	<i>Moreover, since the first half of the twentieth century, the types of translation texts have been diversified and the translation phenomenon has become localized with technological developments</i>

In the example 3, the –le suffix in the word *gelişmeler-le*, means “with” in English. However, it was translated as if it means “in” in the MT. It could be translated as “with technological developments” or “through technological developments”, so that the meaning of “instrument” could be inferred. This error was classified under syntactic errors for being an error in using prepositions.

Example 4: A Semantic Error in the Informative Text

ST	<i>Betimleyici çeviribilim çalışmaları</i>
MT	<i>descriptive transcription studies</i>
Revision	<i>descriptive translation studies</i>

In Example 4, the term *çeviribilim* [literally “science of rendition”] can be interpreted in two different ways. On the one hand, it can refer to a transcription, a written or printed version of something. On the other hand, it can also mean translation studies, the process of translating words or text from one language into another. Judging from the overall meaning of the context, the correct word here should be the latter, even though the term is rendered as the former in MT output. This error was classified under semantic errors for being an instance concerning erroneous choices for polysemes.

Example 5: A Pragmatic Error in the Informative Text

ST	<i>Eylemi yönlendiren ve sonucuna ulaşmasında belirleyici rol oynayan etmen “amaç”tır. Bu kuramın merkezinde yer alan “amaç”, kuramın adını da belirlemiştir. “Skopos” Yunanca kökenli bir sözcük olup, “amaç”, “sonuç” ve “işlev” anlamına gelmektedir.</i>
MT	<i>The factor that directs action and plays a decisive role in reaching the result is the "goal". The "aim" at the center of this theory also determines the name of the theory. "Skopos" is a Greek word, meaning "purpose", "result" and "function".</i>
Revision	<i>The factor that directs action and plays a decisive role in reaching the result is the "aim". The "aim", which is at the center of this theory, also determines the name of the theory. "Skopos" is a Greek word, meaning "aim", "result" and "function".</i>

In the example above, there is a network in the ST established with the words *amaç* [aim]. Because of the fact that the name of the theory introduced in the text is “Skopos” meaning “aim”, the author must have chosen to use the word “aim” in describing other features of the theory in the ST. The network of the author strengthened the meaning in the ST whereas the use of synonyms of aim in the MT weakened the meaning. This error was classified under pragmatic errors for being an instance concerning the destruction of the network of the ST, which is constructed by a pragmatic strategy.

3.2. Analysis of expressive text

Among the 26 total instances of translation errors in the expressive text, 4 instances were defined as Lexical Errors, 2 as Syntactic Errors, 10 as Semantic Errors, and 10 as Pragmatic Errors. One example that has been chosen from each error types below is illustrative.

Example 6: A Lexical Error in the Expressive Text

ST	Ana sütü gibi tath
MT	Sweet as main milk
Revision	Sweet as mother's milk

In the Example 6, the word “ana sütü” means the milk obtained from the mothers. It was translated as “main milk”, since the term “ana” literary means both “main” in the meaning of “essential” and “mother”. This error was classified under lexical errors since it presents the wrong choice of terminology.

Example 7: A Syntactic Error in the Expressive Text

ST	Her kelimedede bir kat daha artacaksın
MT	Each word will increase one more
Revision	You will increase one more at one word

In the Example 7, the subject of the ST was “you” whereas the subject of the MT became “word”. Since the meaning of the ST was totally altered in the translation due to syntaxes differences between the ST and the MT, the error was classified under syntactic errors.

Example 8: A Semantic Error in the Expressive Text

ST	Birisi ana dilin
MT	Someone in your native language
Revision	One in your mother tongue

In the example above, the ST presents a line that is meaningful with the previous line which is “you will know at least three languages”. The lines intended to carry the meaning of “from these three languages one is your mother tongue”, however in the MT, it can be inferred that the meaning altered to a person saying “someone” instead of “one of these languages” or “one”. This error was classified under semantic errors since it failed to signify the same signified in the source line.

Example 9: A Pragmatic Error in the Expressive Text

ST	Ana avrat dümdüz gideceksin
MT	You will go straight to the main avrat
Revision	You will swear a blue streak/to swear like a trooper

The example above presents a pragmatic error which stems from the usage of the literal translation of a convention-related phrase. The idiom of “ana avrat düz gitmek” [*literal meaning: go straight to mother and wife*] means to swear someone and his/her relatives beginning with his/her mother and wife (Tdk.org.tr). The idiom was translated with a literal and word-for-word strategy, instead of using the equivalent of the idiom in the English speaking cultures.

3.3. Analysis of expressive text

Among the 2 total instances of translation errors in the operative text, 1 instance was defined as Lexical Error while 1 was defined as Pragmatic Error. The following examples illustrate these errors.

Example 10: *The Lexical Error in the Operative Text*

ST	<i>Biz çevirmezsek dünya dönmez!</i>
MT	<i>If we do not turn the world does not return!</i>

The lexical error presented in the Example 10 contains a literal translation. The word “çevirmek” in the ST has multiple Turkish equivalents as “to turn” and “to translate”. The slogan actually has this multiple meaning in itself as a pun. It can be stated that the choice of the MT do not feature the pun in the ST, therefore this choice is regarded as a wrong terminology to use in translation of the slogan.

Example 11: *The Pragmatic Error in the Operative Text*

ST	<i>Biz çevirmezsek dünya dönmez!</i>
MT	<i>If we do not turn the world does not return!</i>

In the Example 11, the translation of the whole slogan was regarded as a pragmatic error, since the function of the ST was ignored by the MT. As a slogan, the ST contains a dialogism by using the term “çevirmek” for both to mean “to rotate the World” and “to translate” in the World. The function of the ST was creating a pun over to translate, whereas the function of the translation was altered to just the job of rotating the World.

To sum up, the data analysis revealed that total 66 translation errors were found in Google Translation of an informative text, an expressive and an operative text translated from Turkish to English. According to the statistical test, the text type which presents the most number of translation errors was determined as operative text with 50% translation errors of all lexical items in this text; it was followed by expressive text (15.95%). Informative text recorded relatively lower percentages (4.2%). Due to the plain structure of informative texts, it is not surprising to reach more accurate translations in this type rather than the translations of expressive and operative texts which comprise semantic and pragmatic errors due to their needs for creativity.

Moreover, the data analysis revealed that translation errors varied by text types. It is observed that the error patterns between the text types differed; while the informative text mainly includes *lexical errors*, others mainly include *semantic and pragmatic errors*.

The results of this study may indicate that *Google Translate*'s most frequently occurring errors can be found the expressive and operative texts. The most common errors found in these texts are semantic, pragmatic, and lexical errors which may firstly considered as resulting from nonequivalence between the source (Turkish) and the target language (English), secondly they may result from the lack of corpus of machine translation. This study also revealed that *Google Translate* tend to deliver word-for-word translations, but not sense-for-sense translations. However, even with a word-for-word translation, a number of mistakes can still be found especially for words that have multiple meanings and functions.

As it has been discussed in the introductory section of the study, agreed on Aiken and Balan (2011) suggesting *Google Translate* translates a European language into another European language much

better than other languages, the occurrences of errors are found to be pertinent to language features of the source language. For being an agglutinative language, Turkish may be the reason for syntactic and morphological errors in translation.

4. Conclusion

This study, analyzing the translation errors in Turkish-to-English translations of *Google Translate*, has provided a piece of evidence contributing to machine translation field in error classification for the Turkish-English language pair and in common errors made by the MT. Within the scope of this study, three different text types have been translated using SMT *Google Translation*, and analyzed in terms of five linguistic error types. As a result, in accordance with the research questions, the data analysis revealed that the main text type that has more translation errors is the translation of *operative text* and *expressive text*, respectively. It is also observed that the error pattern between the text types was different. The informative text mainly includes *lexical errors*, whereas operative and expressive mainly include *semantic* and *pragmatic errors*. Summing up the results, it can be concluded that although *Google Translate* provides much quicker translations among a large number of languages, there is still need for human assistance.

The following limitations of the study should be mentioned: first, the study is based on only Turkish-to-English translation, which may change the accuracy of the translation if compared with English-to-Turkish translation; second, the study used the data acquired from *the statistical machine translation* version of *Google Translate*, any future researches can be conducted by using *neural machine translation* version of *Google Translate*. Therefore, all the findings should be interpreted with caution and need further validation.

References

- Aiken, M., & Balan, S. (2011). An analysis of Google Translate accuracy. *Translation Journal*, 16(2). Retrieved May 23, 2018 from <http://translationjournal.net/journal/56google.htm>
- ÇevBir Slogan. (n.d). Retrieved May 22, 2019 from <https://cevbir.org.tr/genel/cevbirin-kurulusuna-amaclarina-ve-faaliyetlerine-dair>
- Çeviribilim Bir Giriş. (n.d). Retrieved May 22, 2019 from http://ceviribilim.com/?page_id=1496
- Chomsky, N. (1957) *Syntactic Structures*. The Hague: Mouton.
- Google Translate. (n.d.) Retrieved May 22, 2019 from <https://translate.google.com/>
- Gu, K., Ng, H. K., Tang, M. L., & Schucany, W. R. (2008). Testing the Ratio of Two Poisson Rates. *Biometrical Journal*, 50(2), 283-298. doi:10.1002/bimj.200710403
- Hernandez, M. S. (2011). Raising students awareness about grammatical and lexical errors via email. *Revista de Lenguas Modernas*, 4, 263-281.
- James, C. (1998). *Errors in language learning and use: Exploring error analysis*. New York: Longman.
- 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.
- Keshavarz, M. H. (1999). *Contrastive analysis and error analysis*. Tehran: Rahnama Publication.
- Munday, J. (2008). *Introducing translation studies: theories and applications* (2nd ed.). Newyork and London: Routledge.
- Quah, C. K. (2006). *Translation and technology*. Basingstoke: Palgrave Macmillan.

Reiss, K., & Vermeer, H. J. (1971, 2013). *Towards a general theory of translational action: skopos theory explained*. (C. Nord, Trans.). Manchester, UK: St. Jerome Publishing.

Sager, J.C. (1994). *Language Engineering and Translation: Consequences of Automation*. Amsterdam: John Benjamins.

Todd, L. (1987). *An Introduction to Linguistics*. Essex: England Person Education Limited.

Yurdakul, M. (2007), *Türkçe Şiirler* (Ed. Hasan Kolcu - Fatih Kıran), İstanbul: Çağrı.