International Journal of English Language and Linguistics Research
Vol.5, No.2, pp.15-23, April 2017
Published by European Centre for Research Training and Development UK (www.eajournals.org)

ANALYZING INDONESIAN-ENGLISH ABSTRACTS TRANSLATION IN VIEW OF TRANSLATION ERRORS BY GOOGLE TRANSLATE

Sependi Napitupulu
Department of English Literature, Faculty of Letters
Methodist University of Indonesia, Medan

ABSTRACT: *This study seeks to investigate the frequency of errors in the translation of abstracts produced by Google Translate with reference to Keshavarz's (1999) model of error analysis. This research will be of great benefit to undergraduate students to use these findings as a guideline in writing a thesis abstract. Five types of error classification is used as the parameters, namely lexicosemantic error, tense error, preposition error, word order error, distribution and use of verb group error, and active and passive voice error. The data were obtained from several faculties at the Methodist University of Indonesia, Medan. A total of ten abstracts of undergraduate students' paper from various faculties were randomly selected. The data are then compared on each sentence segment and any words or phrases found to have errors are analyzed. The study revealed that 21 frequencies in terms of lexicosemantic errors, 9 frequencies in terms of tense errors, 13 frequencies in terms of preposition error, 27 frequencies in terms of word order error, 15 frequencies in terms of distribution and use of verb group errors, 8 frequencies in terms of active and passive voice errors.*

KEYWORDS: Error Analysis, Translation Errors, Abstract Translation, Machine Translation

INTRODUCTION

The use of Google Translate has been increasing either in the academic discipline or in the non-academic discipline. Despite the fast-turnaround time produced by Machine Translation such as Google Translate, the quality of the translation has been considered far from perfection. Thus in order to evaluate the quality of machine translation, error analysis has been suggested to be conducted. In line with this, numerous text genres have been investigated, including one of which is an abstract text. Abstracts as a summary of a research paper harbors important information where it serves to attract readers to whether read the entire passage or leave it.

In Indonesia, the need to translate abstracts of undergraduate thesis into English is a requirement for students to complete their final year academic papers. Some universities oblige students to have the abstracts of their scientific papers translated by an official institutional language center, but, some others do not require official translation service to translate their abstracts into English. As a result, students are allowed to use whatever means available to translate their abstracts as fast as possible without taking into account the accuracy of the translated text. The last resort students would take is using Google Translate to translate their abstract since the service it provides is free and real-time basis results within seconds. However, the majority of students have been unaware of the consequences it bears upon it such as wrong word choice, wrong prepositions, wrong word order, and so on. They have been known to have used Google Translate to translate their abstracts of bachelor's papers into English. This phenomenon is found in the Methodist University of Indonesia, Medan. The university administration does not require final year students to translate their bachelor's papers' abstracts accurately using the service of professional translator.

ISSN 2053-6305(Print), ISSN 2053-6313(online)

Purpose of the Study
The main purpose of this study is to find out the errors created by Google Translate in translating abstracts of bachelor's papers from Indonesian to English.

Research Question
What is the frequency of errors in the translation of abstracts form Indonesian into English in terms of errors in lexicosemantic, tense, preposition, word order, distribution and use of verb group as well as active and passive voice?

LITERATURE REVIEW

Translation
House, J. (2015:2) states that translation can be defined as the result of a linguistic-textual operation in which a text in one language is re-contextualized in another language. As a linguistic-textual operation, translation is, however, subject to, and substantially influenced by, a variety of extra-linguistic factors and conditions.

Machine Translation
Google Translate is one of several machine translations most commonly used by people around the world to translate texts over 90 different languages. Not only can it translate words, but also phrases, sections of a text, or a web page. To translate a text, Google Translate searches different documentaries to find the best appropriate translation pattern between translated texts by human. This pattern searching is called Statistical Machine Translation. Since the number of translated texts varies from users to users, consequently, the quality of Google Translate depends on the number of human translated texts searched by Google Translate (Karami, 2014). Quite recently, another assessment to the study of Google Translate has been proposed by Bozorgian and Azadmanesh (2015). In case of subject-verb agreement, they considered both Google Translate and human translators and finally they concluded that Google Translate does not handle subject-verb agreement very well while translating English sentences into Persian compared to human translators.

In line with present research, several studies have done similar researches on abstract translation errors. Setiawan (2013) has done a research on English translation errors in abstracts of educational administration students of post graduate school of State University of Medan. He classifies the errors into two types namely grammatical errors and content errors. His research found that in the grammatical error type, the use of verb group took 58%, and (2) in the content error type, presentation of different senses took 38%.

In the meantime, Susilowati (2006) conducted a research on translation errors in the English-translated abstracts of some management department theses in Petra Christian University. Her study discovered that from the quantity of the content errors in translation, the majority of Management Department students make left out and different sense errors. For the grammatical error, the predominant errors are error in the production of verb group and miscellaneous error. Content errors include different senses, less precise or less accurate, misuse of phrases, misuse of formal or official registers, etc. (Newmark, 1988; Barnett and Stubbs, 1980). Corder (1974) was the first who studied error analysis and defined language transfer as the main process in L1/L2 language learning in the 1960s. Meanwhile, Keshavarz (1999) defined error analysis as collecting samples, identifying errors, classifying, and evaluating them. Error analysis, as a

ISSN 2053-6305(Print), ISSN 2053-6313(online)

way to identify systems' weaknesses and define priorities for their improvement, is gaining increasing interest in the MT community (Popovic and Ney, 2011; Popovic et al., 2013).

Error Analysis

Error Analysis (EA), a fundamental branch of applied linguistics, emerged in the sixties to address students' performance (Shrestha, 1979). According to Longman Dictionary of Language Teaching & Applied Linguistics (2010), (EA) is manifested in order to (1) describe strategies used by the learners in language teaching, (2) spot causes of errors, and finally (3) gain information on common difficulties in language learning to develop materials and strategies to help the learners avoid their errors.

Error analysis is a type of linguistic analysis that focuses on the errors learners make. It consists of a comparison between the errors made in the Target Language (TL) and that TL itself (Corder, 1974). According to Richards et al. (1992: 96), "error analysis may be carried out in order to: a) find out how well someone knows the language, b) find out how a person learns a language, and c) obtain information on common difficulties in language learning". Moreover, error analysis explores analytically the actual errors which are produced by foreign language learners and tries to describe the causes of errors. Moreover, error analysis helps to identify the weaknesses, with a variety of techniques, for identifying, classifying and systematically interpreting the language learners' errors (Khodabandeh, 2007).

In terms of errors made by students of English, errors appear when the learner's knowledge of the rules of the target language is incomplete. Errors are considered to be systematic, governed by rule and also regarded as rule-governed when they follow the rules of the learner's interlanguage (Keshavarz, 2011). According to Abbasi and Karimnia (2011) it is essential that teachers be able to adjust their teaching plan to make their teaching work more effectively by identifying learners' errors. Moreover, recognizing errors can provide valuable information for teachers about how much the learner has learned and what kind of problems s/he has in the study of language. As Conde (2011) puts it, "error detection has been the traditional basis for translation evaluation". Gass and Selinker (1994) identified six steps followed in conducting an error analysis. These included 'collecting data', 'identifying errors', 'classifying errors', 'quantifying errors', 'analyzing sources of error', and 'remediating for errors.'

Errors in Translation

According to Koller (1979:216), based on the concept of equivalence between source text and target text, a translation error is regarded as some kind of non-equivalence between source text and target text or non-adequacy of the target text. In functionalistic approach and approaches based on the 'skopos theory', an error is defined as relative to the fulfillment of the target text function and the receiver's expectations (Schmitt 1998:394; Nord 2009; 190).

More recently, error analysis has also become a focus of research in the Machine Translation (MT) area, where some works are dedicated to the design of taxonomies (Llitj´os et al, 2005; Vilar et al, 2006; Bojar, 2011) and others target errors' identification (Popovi´c and Ney, 2006). Evaluation of machine translation (MT) ouput is a controversial task in the MT community. Several automatic measures have been proposed including the Word Error Rate (WER), the Position independent word Error Rate (PER), the BLEU (Papineni et al., 2002) and the NIST (Doddington, 2002) measures being the most widely used ones. A relationship between these error measures and the actual errors found in the translations is however not easy to find. Since

ISSN 2053-6305(Print), ISSN 2053-6313(online)

present study is a qualitative study, and those measures are of qualitative study, then they are used in the present study. We adopt error classification based on grammatical errors, lexical errors, word order errors which are easier to investigate and those error classification further elaborated on section three below.

METHODS

This study applies a descriptive qualitative approach. In collecting the data, a total of 10 abstracts were randomly selected from various faculties including the faculty of agriculture, the faculty of letters, the faculty of information and technology. All abstracts were in fact translated using machine translate, that is Google Translate. From the ten abstracts compiled, a total of 92 sentence fragments were gathered.

In analyzing the collected data, the first step taken was tabulating and comparing the frequencies of Indonesian to English translation errors of different types. Then, the frequencies of correct and incorrect translated tokens of the different types of translation errors according to Keshavarz's (1999) model of error analysis (e.g., lexicosemantic errors, tense errors, wrong use of preposition, word order errors, errors in the distribution and use of verb groups, and errors related to active and passive voice) were put together in separate tables. Then, the frequencies of different types of errors produced by Google Translate were counted, tabulated, and the proportion of each type was converted into percentages.

RESULTS AND DISCUSSION

Below are some examples of errors found in the translation of abstracts

Error in Lexicosemantic

Shown below is an example of a text that has errors in lexicosemantic.

Source Text	: Luaran dari penelitian ini berupa software untuk mencari dan **menghapus file yang di duplikat** pada media penyimpanan computer.
GT Text	: Outcomes of this research is a software to find and **remove duplicate files** on computer storage.
Suggested Translation	: The outcome of this study is to produce a software that can search for and delete duplicated files on a computer storage device.

As we can see, GT failed to translate the clause "*menghapus file yang di duplikat*" in Bahasa, instead it was translated into "remove duplicate files". It should have been "remove duplicated files". GT has been known to produce inappropriate modifier and modified rule.

Tense

As the publication manual of the American Psychological Association claims (2002, p. 14), "use the present tense to describe results with continuing applicability or conclusions drawn; use the past tense to describe specific variables manipulated or test applied". Generally, students used present tense in their abstracts. They were mixing present and past tenses. This also is found in machine translation system, as shown in the below example of the text that has errors in tense.

Source Text	: Data diambil dari teknik novel *The Grapes of Warth* dan dari informan (petani) di daerah Tanah Karo, Raya, Tiga Tunggu, dan Saribudolok.

ISSN 2053-6305(Print), ISSN 2053-6313(online)

GT Text	: Data **are** taken from *The Grapes of Warth* and the Informants (farmers) in tanah Karo, Raya, Tiga Runggu, and Saribu dolok.
Suggested Translation	: Data were collected from *The Grapes of Warth* as well as from some informants (farmers) in several districts including Tanah Karo, Raya, Tiga Runggu, and Saribu dolok.

As pointed above, the use of past tense is typical in writing abstract. However, in Indonesian grammar, there is no tense, but in the English grammar, tense has substantial role in emphasizing the order of an event. The verb "are" in the above example of GT text should have been "were" since the data has been collected in the past.

Preposition

Shown below is an example of the text that has errors in preposition.

1.	Source Text	: **pada** media penyimpanan computer secara berulang-ulang.
	GT Text	: the computer storage media over and over again.
	Suggested Translation	: on the computer storage media repeatedly.
2.	Source Text	: *Dilaksanakan di kampus Fakultas Pertanian UMI **sekitar** 13 bulan (Mei 2013-Juni 2014).*
	GT Text	: Held on the campus of the Faculty of Agriculture UMI **about** 13 month (may 2013-June 2014).
	Suggested Translation	: The research was conducted at the Faculty of Agriculture, UMI campus for 13 month (may 2013-June 2014).

As is presented above, the Indonesian text starts with a preposition "pada", but in the Google Translated Text, the preposition is omitted or left untranslated. In this case, the text sounds unfaithful as the original text. As for the second text, the preposition "sekitar" in Indonesian is translated by Google Translate into "about". In this sense, GT failed to adopt the context, in which the original text actually means "duration" of the research taking place.

Word Order

Word order is the one of the most common types of errors produced by google translate. Shown below is an example of the text that has errors in word order.

Source Text	: mencopy file dari media penyimpanan yang lain yang nama filenya berbeda tetapi contenct dari file nya yang sama.
GT Text	: copy files from storage media whose file names are different **but the content of the same file.**
Suggested Translation	: copying files from other storage media whose file names are different but the content of the file is similar**.**

The above sentence shows that Google Translate does not translate the source clause "tetapi contenct dari file nya yang sama" properly, instead it was translated as "file names are different but the content of the same file". The original meaning in the source text is actually "the

ISSN 2053-6305(Print), ISSN 2053-6313(online)

content" being similar not "the file" being similar. Therefore the original meaning of the text is not successfully transferred and it caused semantic error as well.

Distribution and Use of Verb Group

In this classification of distribution and use of verb group error, this study refers to Ho's (2005) error taxonomy regarding verbs and verb groups:

a. Omission of suffix – s/- es/ - ed/ - ing
b. Omission of – ed participle after a form of the verb be
c. Inappropriate form after modal verb (unnecessary insertion/ overgeneralization of suffix – s, - ed, - ing, infinitive to)
d. Inappropriate form after modal verb (omission of suffix – ed in past perfect form)
e. Omission of direct object
f. Omission of infinitive to
g. Omission of –ing participle
h. Omission of auxiliary verb
i. Inversion of verb - subject in indirect question format

In addition to the above classification, this research also looks at the wrong use of verb forms in sentences. Below is an example of wrong verb form in the target text produced by Google Translate.

Source Text	: Prototipe penyirma tanaman **menggunakan** sensor suhu dan kelembapan berbasis mikrokontroler arduino uno untuk menggerakan secara automatisasi alat selama kelembapan tanah belum tercapai
GT Text	: Penyirma prototype plants **using** temperature and humidity sensors based microcontroller arduino uno to move in automation tools for soil moisture has not been reached.
Suggested Translation	: Plant watering prototype uses microcontroller-arduino-uno-based temperature and humidity sensor to move the tools automatically as long as the soil moisture has not been reached.

As can be seen from the above text, the source word "menggunakan" is translated as "using" in the google translated text. The use of "verb-ing" in the above target text is inappropriate. The form of "verb-ing" in English normally serves as an adjective clause which acts as a modifier.

Active and Passive Voice

Some errors committed by Google Translate is also sometimes the structure of active and passive voice as shown below:

Source Text	: Telah dirancang sebuah alat yang berfungsi untuk memelihara tanaman agar bertumbuh dengan baik
GT Text	: Has designed a tool that serves to maintain the plant to grow well.
Suggested Translation	: A tool has been designed to support the plant to grow well

The above text shows that GT translates the source text in a form of active voice while sentence structure of the source text is in the form of passive voice. Therefore, in this case, the issue here lies in the error classification of passive and active voice. In addition to that, a subject of the

ISSN 2053-6305(Print), ISSN 2053-6313(online)

sentence is missing, therefore, the sentence should sound "A tool has been designed to support the plant to grow well."

Accordingly, the following six types of errors as per Table 1 were identified, counted, and categorized.

Table 1. Frequencies of Errors in the Translation of Indonesian-to-English Abstracts by Google Translate

Error Types	Frequency	Percentage
Lexicosemantic	21	29
Tense	10	14
Preposition	9	12.5
Word Order	15	21
Distribution and Use of Verb Group	9	12.5
Active and Passive Voice	8	11
Total	72	100%

By looking at the second column of the above table, lexicosemantic errors had the highest frequency (f = 21), whereas wrong use of active and passive voice had the lowest frequency (f = 8). Error types which were in the middle were wrong distribution and use of verb group (f = 9), errors relating to preposition (f = 10), word order errors (f = 15), and wrong use of prepositions (f = 9). The total number of identified errors was 72 errors.

The direction of translation might affect the quality of the translations rendered by the Google Translate since the frequencies of errors of each type were mostly different in Indonesian to English renderings from the time the translation was done from English to Indonesian.

CONCLUSION

This study has investigated types of errors found in the translation of Indonesian bachelor's papers of the students of Methodist University of Indonesia, Medan into English as a result of relying Google Translate as the tool on translating the text. Lexicosemantic as the dominant error found in abstract translations has affected the content of the abstract. Some words have been mistranslated and do not fit the context of the original text. Therefore, it is encouraged to alter the words done by human translators by looking at the context as a whole.

As for academics, it is suggested that further researches touch on similar issues in abstract translation but with more varieties from different universities or colleges with a larger amount of data and different backgrounds or fields of study.

REFERENCES

Abbasi, M., Karimnia, A. 2011. "An Analysis of Grammatical Errors among Iranian Translation Students: Insights from Interlanguage Theory". European Journal of Social Sciences. , Vol.25 No.4 pp. 525-536. 2011. Retrieved November 23, 2012. [http://www.europeanjournalofsocialsciences.com]Doddington, 2002

Barnett, S & Marcia S. (1980). Practical Guide to Writing. Canada: Little, Brown & Company.

Bojar O (2011) Analysing Error Types in English-Czech Machine Translation. The Prague Bulletin of Mathematical Linguistics pp 63–76.

ISSN 2053-6305(Print), ISSN 2053-6313(online)

Bozorgian and Azadmanesh (2015). A survey on the subject-verb agreement in Google Machine Translation. International Journal of Research Studies in Educational Technology, 4 (1), April 2015, pp. 51–62.

Conde. T. (2011). Translation evaluation on the surface of texts: a preliminary analysis. Conde, Departament de Traducció i Comunicació, Universitat Jaume I. Jostrans. The Journal of Specialized Translation. Issue15. Accessed from http://www.jostrans.org/issue15/art_conde.php

Corder, S.P. (1974) Error Analysis. In J. P. B. Allen and S. Pit Corder (eds.) Techniques in Applied Linguistics (The Edinburgh Course in Applied Linguistics:3), London: Oxford University Press (Language and Language Learning), pp. 122-154.

Costa. A. et. al. A Linguistically Motivated Taxonomy for Machine Translation Error Analysis

Gass, S. & Selinker, L. (1994). Second language acquisition: An introductory course. Hillsdale, NJ: Erlbaum. Genoa, Italy, pp 697–702

Ghasemi. H., Hasemian. M. (2016). A Comparative Study of Google Translate Translations: An Error Analysis of English-to-Persian and Persian-to-English Translations. English Language Teaching 9(3):13. January 2016. DOI: 10.5539/elt.v9n3p13

Hansen, G. 2010. Translation 'errors'. In: Gambier, Y., van Doorslaer, L. (eds). Handbook of Translation Studies. Amsterdam: John Benjamins. 385–388.

Ho, M. L. (2005). Exploring errors in grammar: A guide for English language teachers. Singapore: Pearson Education. South Asia Pte Ltd.

House, J. (2015). *Translation Quality Assessments*. Routledge

Karami, O. (2014, January). 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, 2011). *Contrastive analysis and error analysis.* Tehran: Rahnama Publication.

Khodabandeh, F. (2007). Analysis of students' errors: The case of headlines. *The Asian ESP Journal, 3*(1), 6-21.

Kishore Papineni, et al. (2002). Corpus-based comprehensive and diagnostic MT evaluation: Initial Arabic, Chinese, French, and Spanish results. In Proceedings of Human Language Technology 2002, San Diego, CA. To appear

Koller, W. 1989 (1979). Equivalence in Translation Theory. In Andrew Chesterman (ed.), Readings in Translation Theory. Loimaa: Finn Lectura, 99-104.

Llitj´os AF, Carbonell JG, Lavie A (2005) A Framework for Interactive and Automatic Refinement of Transfer-Based Machine Translation. In: In Proceedings of EAMT 10th Annual Conference, pp 30–31

Longman Dictionary of Language Teaching &Applied Linguistics (4th ed.). (2010). London: Pearson Press.

Newmark, P. (1988). A Textbook of Translation. London: Prentice Hall Europe.

Nord. C. (2009). *Text Analyse and Ubersetzen.Theoretische Grundlagen, Methode und didaktische Anwendung einer übersetzungsrelevanten Textanalyse*. Tübingen: Julius Groos Verlag Brigitte Narr. (četrta izdaja).

Popovic, M. et. al. (2013). Learning from human judgments of machine translation output. In Proceedings of the MT Summit XIV. Proceedings of MT Summit XIV.

Popovic. M. and Ney. H. (2011). Towards automatic error analysis of machine translation output. Comput. Linguist., 37(4):657–688, December.

Publication manual of the American Psychological Association. (2002). Fifth edition. USA: APA.

ISSN 2053-6305(Print), ISSN 2053-6313(online)

Richards, Et al. (1992). Dictionary of language teaching and applied linguistics. Essex: Longman.

Schmitt, P. (1998). "Qualitatsmanagement." In Handbuch Translation, Mary Snell-Hornby, Hans G. Honing, Paul Kussmaul & Peter A. Schmitt (eds), 394-399. Tubingen: Narr.

Setiawan, Y. (2013). *English Translation Errors In Abstract Of Educational Administration Students.* Unpublished Master's Thesis, UNIMED.

Shrestha, A. (1979). "Error Analysis: A Pragmatic Approach." CNAS Journal, 6, 2: 1-9.

Susilowati, I. (2006) Translation errors in the English-translated abstracts of some Management Department theses in Petra Christian University. Unpublished Bachelor Thesis, Petra Christian University.

Vilar D, Xu J, D'Haro LF, Ney H (2006) Error Analysis of Machine Translation. Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC-2006), Genoa, Italy. European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2006/pdf/413_pdf.pdf.

Vilar, D. 2006. Error Analysis of Statistical Machine Translation Output. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 06)*, pages 697–702, Genoa, Italy, May.

ISSN 2053-6305(Print), ISSN 2053-6313(online)