Versi Online: http://journal.ubm.ac.id/xxx/xxx
DOI : dx.doi.org/xxx/xxxxxxxx
Hasil Penelitian

Journal of English Language and Culture
Vol. 10 (No. 2) : 64 - 74. Th. 2020
p-ISSN: 2087-8346
e-ISSN: 2597-8896

PRE-EDITING OF GOOGLE NEURAL MACHINE TRANSLATION

Alvin Taufik*

English Language and Culture Department, Bunda Mulia University

Received on 17 February 2020 / Approved on 28 February 2020

ABSTRACT

Even with the new Machine Translation (MT) platform available in Google today (Neural, as compared to the previous Statistical one in the previous years), the output is not always satisfactory. This is even more obvious in specific contexts and situations. Research has shown that the implementation of rules for the process prior to and the one that follows the input activities into an MT (often referred to as the pre-editing and post editing process) has proven to be fruitful (Gerlach, et. al., 2013; Shei, 2002). However, to the best knowledge of the researcher, no research on pre-editing rules on Indonesian input into MT has been conducted. This research is significant because it might increase efficiency and effectiveness of MT, especially for the language pair Indonesian-English. For that reason, this research intends to identify the pre-editing rules required to create a solid basis to translate Indonesian Source Text (ST) into English Target Text (TT). This research adopts the product-oriented research. The results show that in the pre-editing process, the length of the sentence, the conjunctions (subordinative and correlative), and the inappropriate ST words should be the focus of attention.
Keywords: machine, pre-editing, translation

ABSTRAK

Meskipun saat ini Terjemahan Mesin (TM) dengan platform *baru telah tersedia di Google (Neural, dibandingkan dengan yang berbasis Statistik di beberapa tahun yang lalu), hasilnya tidak selalu memuaskan. Fakta tersebut bahkan lebih nyata dalam konteks dan situasi tertentu. Penelitian terdahulu telah menunjukkan bahwa implementasi pedoman untuk proses sebelum dan sesudah input (atau proses pra- dan pasca editing) telah terbukti bermanfaat (Gerlach, et. Al. , 2013; Shei, 2002). Namun, sepengetahuan peneliti, belum ada penelitian tentang pedoman pra-edit untuk input bahasa Indonesia ke TM. Penelitian ini penting karena dapat meningkatkan efisiensi dan efektivitas TM, terutama untuk pasangan bahasa Indonesia-Inggris. Oleh karena itu, penelitian ini bermaksud untuk mengidentifikasi pedoman pra-edit bahasa sumber yang baik untuk di input ke dalam TM. Penelitian ini mengadopsi penelitian yang berorientasi produk. Hasil penelitian menunjukkan bahwa dalam proses pra-edit, penerjemah atau penyunting kemungkinan besar harus menyunting bahasa sumber yang berkaitan denganh panjang kalimat, bentuk-bentuk konjungsi (subordinatif dan korelatif), dan kata-kata yang tidak tepat.*
Kata Kunci: *mesin, pra-edit, penerjemahan*

INTRODUCTION

Hailed as being close to human-level translation, the products of Neural Google Machine Translation (NGMT) actually relates to specific language pairs (Statt, 2006). It is interesting to see whether the language pair of Indonesian-English will be as accurate as others. The research is conducted to provide insights on the alternative solution to the process of translation using MT, which has mostly relates to the process of post-editing. Specifically, this research is conducted to identify the pre-editing rules which can be implemented in translating Indonesian text into English using Neural Google MT.

*Author(s) Correspondence:
E-mail: ataufik@bundamulia.ac.id

Versi Online: http://journal.ubm.ac.id/xxx/xxx
DOI : dx.doi.org/xxx/xxxxxxxx
Hasil Penelitian

Journal of English Language and Culture
Vol. 10 (No. 2) : 64 - 74. Th. 2020
p-ISSN: 2087-8346
e-ISSN: 2597-8896

RELATED WORKS

Japanese researchers Miyata and Fujita (2017) have conducted similar research using the language pair of Japanese to English. They are using four different data sets and off-the-shelf MT system. In addition, they follow the human-in-the-loop protocol. This refers to the inclusion of human pre-editing process in the MT production process. The results of their research show that the pre-editing process enables 85% accuracy of the MT product. Their research also points out that the pre-editing process is benefial in multilingual setting as they also used the pre-edited ST to be translated into Chinese and Korean.

Other jepanese researchers, Hiraoka and Yamada (2019) has identified rules to be used in MT-based translation of Japanese into English. Their rules include the insertion of punctuation, explicitation of implied subject and object, and the writing of proper nouns in English. This is however three of nineteen (19) rules of pre-editing. The three chosen rules are based on frequency, ease of use, and the potential editor (with the editors in focus being non-bilinguals. The evaluation method implied to evaluate the result is BLEU score and human evaluation. From the research it was found that quality was significantly improved on human evaluation.

FRAMEWORK

Pre-editing

There are not that many reference on pre-editing. One of which is by Hiraoka and Yamada (2019), which distinguished pre-editing into two: bilingual and monolingual. The difference between the two of them is that in bilingual pre-editing, the editor is allowed to edit the ST while looking at MT output. Since corpora is used to evaluate the result of the MT, this research is focused on bilingual pre-editing.

Human evaluation

This research adapts human evaluation used by Hiraoka and Yamada (2019). The evaluation is done on the Result of MT before the Pre-edit Process and Result of MT after the Pre-edit Process.

METHODOLOGY

This research focuses on the result of MT on an academic texts. The texts are chosen based on the researcher's experience on translating, which are projected as a potential cases of translation problems. Furthermore, related to the stages in this research, the data is coded into the followings:

ST	Source Text
Raw TT	Result of MT before the Pre-edit Process
PrE ST	Pre-edited Source Text
PrE TT	Result of MT after the Pre-edit Process

To objectively evaluate the quality of the MT outputs, the researcher adopts the Corpus Linguistic-based product-oriented approach. These research is divided into 6 (six) stages. In the first stage, the ST is translated into English using the MT. In the second stage, the Raw TT is compared to the sub BNC corpora. Based on that, at the third stage, analysis is conducted on the ST. This analysis become the basis for the hypothesis to be tested on the final stage. At the fourth stage, the ST is pre-edited to create a PrE ST. Next, the PrE ST is inputted into the MT to create PrE TT. On the next stage, the hypotheses gathered from the analysis are further tested on another set of text. This stage is repeated if further findings, which eventually leads to more hypothesis(es) are observed. Finally, the Raw TT and the PrE TT is evaluated using human evaluation. The human evaluation is conducted with Indonesian native speaker who are also lecturer of English. The human evaluation is conducted with a category modified from Hiraoka and Yamada's model. The

*Author(s) Correspondence:
E-mail: ataufik@bundamulia.ac.id

Versi Online: http://journal.ubm.ac.id/xxx/xxx
DOI : dx.doi.org/xxx/xxxxxxxx
Hasil Penelitian

Journal of English Language and Culture
Vol. 10 (No. 2) : 64 - 74. Th. 2020
p-ISSN: 2087-8346
e-ISSN: 2597-8896

modifications are made based on the findings. The human evaluation evaluation model is as follows:

Table 1. Modified Three-Grade Scale of Translation Evaluation

Score	Naturalness	Readibility
3	It is very natural. Structural naturalness is achieved.	It is very easy to read
2	It is acceptable. Source Language structure is still visible.	It is quite difficult to read. Much efforts are needed.
1	It is not a good translation.	It is not readable.

FINDINGS AND DISCUSSION

The first text is taken from an abstract of an academic article. The following is the ST:

ST

Kolokasi merupakan gabungan kata yang memiliki hubungan makna dan selalu muncul berdampingan. Hubungan makna menyebabkan kolokasi tersebut tidak dapat digantikan dengan kata lain, atau jika digantikan dengan kata lain, akan menyebabkan perubahan makna. Tujuan penelitian ini adalah untuk menjelaskan pembentukan kolokasi dari bidang semantik kognitif. Penelitian ini merupakan penelitian campuran yang mengkombinasikan penelitian kuantitif dan kualitatif. Objek penelitian adalah kolokasi verba dan objek bahasa Mandarin bidang transportasi yang bersumber dari korpus data Chinese Web 2017 (zhTenTen11) Simplified Sketch Engine. Hasil penelitian terhadap 900.029 kolokasi verba dan objek bidang transportasi menunjukkan bahwa konsep yang direpresentasikan oleh karakter Han berperan dalam pembentukan kolokasi, komponen makna dari verba berasosiasi dengan komponen makna dari nomina membangun makna kolokatif menyebabkan kolokasi tersebut tidak tergantikan, bingkai menunjukkan perbedaan perspektif dan interaksi elemen pembentuk kolokasi, dan pemaduan konseptual terbukti dapat digunakan sebagai alat untuk menggambarkan dan menjelaskan pembentukan kolokasi verba dan objek. Di samping itu, hasil penelitian juga menunjukkan bahwa kolokasi verba dan objek bahasa Mandarin mencerminkan pemahaman penutur bahasa Mandarin yang memahami kejadian berdasarkan interaksi anggota tubuh dengan benda secara keseluruhan dan detail.

And below is the TT:

Collocation is a combination of words that have a ***relationship of meaning*** *and always appear side by side.* ***The meaning relationship causes the collocation cannot be replaced with other words, or if replaced with other words, will cause changes in meaning.*** *The purpose of this study is to explain the formation of collocation from the cognitive semantic field. This research is a* ***mixed study*** *that combines quantitative and qualitative research. The object of research is the collocation of verbs and Chinese objects in the field of transportation sourced from the Chinese Web 2017 data corpus (zhTenTen11) Simplified Sketch Engine. The results of the study of 900,029 collocation of verbs and objects in the field of transportation show that the concepts represented by Han characters play a role in the formation of collocations,* ***the meaning components of verbs associated with the meaning components of nouns construct the collocative meaning, causing the collocation to be irreplaceable, the frame shows differences in perspective and the interaction of the forming elements collocation and conceptual integration are proven to be used as tools to describe and explain the formation of collocation of verbs and objects****. In addition, the results of the study also showed that the collocation of verbs and objects of the Chinese language reflected the understanding of Mandarin speakers who understood events based on the interaction of the limbs with objects as a whole and in detail.*

*Author(s) Correspondence:
E-mail: ataufik@bundamulia.ac.id

Versi Online: http://journal.ubm.ac.id/xxx/xxx
DOI : dx.doi.org/xxx/xxxxxxxx
Hasil Penelitian

Journal of English Language and Culture
Vol. 10 (No. 2) : 64 - 74. Th. 2020
p-ISSN: 2087-8346
e-ISSN: 2597-8896

Immediately, there are two phrases which appears to be unnatural initially, they are 'relationship of meaning' and 'mixed study'. Both phrases originate from English which has been translated into Indonesian. A quick look at the corpora, it is confirmed that the word relation does not co-occur with meaning in academic setting. So, the researcher decided to search the web for the proper collocation and found the collocative phrase 'sense relation'. This new phrase is tested on the corpora and it yields positive result.

For the phrase 'mixed study', after being tested on the corpora, it turns out that the phrase is used in academic context. From this early finding, the early hypothesis to be made is as follows:

H1:
For academic terminologies, pre-editing is unnecessary since the common ones (such as penelitian campuran) can be successfully translated by MT, while the field-specific ones (such as hubungan makna) needs more post-editing than pre-editing.

For the second case, each highlighted sentence has different problem. In the first sentence, it tends to use exactly the same structure as the Indonesian structure and writing style. As for the second sentence, the excessively long sentence become the obvious problem. In my opinion, the MT result which follows the structure of ST needs more post-editing than pre-editing. So, this example will not be considered in making the pre-editing rules. The long sentence, however, can be made better, and the process can be considered to be included in the rule.

Looking at the sentence, it is immediately obvious that it can be divided into two sentences, with the first one being '*Hasil penelitian terhadap 900.029 kolokasi verba dan objek bidang transportasi menunjukkan bahwa konsep yang direpresentasikan oleh karakter Han berperan dalam pembentukan kolokasi, komponen makna dari verba berasosiasi dengan komponen makna dari nomina membangun makna kolokatif menyebabkan kolokasi tersebut tidak tergantikan',* and the rest as the second sentence. So, in the pre-editing stage, that sentence is divided into two sentences.

Moreover, in the first sentence, between the words 'kolokasi' and 'komponen', and the words 'verba' and 'berasosiasi', the researcher belives it can be made better by adding conjunction. To relate the complex sentence, the researcher included subordinating conjunction of 'dimana' and 'yang' respectively.

From this finding, the second hypothesis can be made:

H2
For academic texts pre-editing in the form of sentence division is necessary on longer sentences, and there is a need to include conjunction in such sentence.

After the pre-editing stage, the following is the PrE ST:

PrE ST
Kolokasi merupakan gabungan kata yang memiliki hubungan makna dan selalu muncul berdampingan. Hubungan makna menyebabkan kolokasi tersebut tidak dapat digantikan dengan kata lain, atau jika digantikan dengan kata lain, akan menyebabkan perubahan makna. Tujuan penelitian ini adalah untuk menjelaskan pembentukan kolokasi dari bidang semantik kognitif. Penelitian ini merupakan penelitian campuran yang mengkombinasikan penelitian kuantitif dan kualitatif. Objek penelitian adalah kolokasi verba dan objek bahasa Mandarin bidang transportasi yang bersumber dari korpus data Chinese Web 2017 (zhTenTen11) Simplified Sketch Engine. Hasil penelitian terhadap 900.029 kolokasi verba dan objek bidang transportasi menunjukkan bahwa konsep yang direpresentasikan oleh karakter Han berperan dalam pembentukan kolokasi dimana komponen makna dari verba yang berasosiasi dengan komponen makna dari nomina membangun makna kolokatif menyebabkan kolokasi tersebut tidak tergantikan. Bingkai menunjukkan perbedaan

*Author(s) Correspondence:
E-mail: ataufik@bundamulia.ac.id

Versi Online: http://journal.ubm.ac.id/xxx/xxx
DOI : dx.doi.org/xxx/xxxxxxxx
Hasil Penelitian

Journal of English Language and Culture
Vol. 10 (No. 2) : 64 - 74. Th. 2020
p-ISSN: 2087-8346
e-ISSN: 2597-8896

perspektif dan interaksi elemen pembentuk kolokasi, dan pemaduan konseptual terbukti dapat digunakan sebagai alat untuk menggambarkan dan menjelaskan pembentukan kolokasi verba dan objek. Di samping itu, hasil penelitian juga menunjukkan bahwa kolokasi verba dan objek bahasa Mandarin mencerminkan pemahaman penutur bahasa Mandarin yang memahami kejadian berdasarkan interaksi anggota tubuh dengan benda secara keseluruhan dan detail.

And this is the PrE TT:

Collocation is a combination of words that have a relationship of meaning and always appear side by side. The meaning relationship causes the collocation can not be replaced with other words, or if replaced with other words, will cause changes in meaning. The purpose of this study is to explain the formation of collocation from the cognitive semantic field. This research is a mixed study that combines quantitative and qualitative research. The object of research is the collocation of verbs and Chinese objects in the field of transportation sourced from the Chinese Web 2017 data corpus (zhTenTen11) Simplified Sketch Engine. The results of the study of 900,029 collocation of verbs and objects in the field of transportation show that the concepts represented by Han characters play a role in the formation of collocations where the meaning component of the verb associated with the meaning component of nouns builds a colocative meaning causing the collocation to be irreplaceable. The frame shows the different perspectives and interactions of the elements forming collocation, and conceptual integration is proven to be used as a tool to describe and explain the formation of collocation of verbs and objects. In addition, the results of the study also showed that the collocation of verbs and objects of the Chinese language reflected the understanding of Mandarin speakers who understood events based on the interaction of the limbs with objects as a whole and in detail.

Having established the hypothesis, the researcher continues with the human evaluation. The result of human evaluation is as follow:

Table 2. Result of Human Evaluation for Academic Text of Raw TT and PrE TT

Raters / Texts	T1 (on Naturalness and Readability)	T2 (on Naturalness and Readability)
R1	2	3
R2	2	3
R3	1 (N) and 2 (R)	2
R4	2	3
Average	**1.875**	**2.75**

Moreover, the raters provide feedbacks on the results of the translation. The followings are the summary of the feedbacks given:

Table 3. Raters' Feedbacks

Raters	Feedbacks
R1	**On the result of the translation:** On further interview, the T1 is rated 2, however, she further explained that the SL structure is actually more visible in T2 (which is rated 3), she gave the explanation of the conjunction 'where' which was actually one of the words included in the pre-edit process. She further added that T2 is more readable and easily understood. **On the rubric:** The rater had difficulties deciding the naturalness and readibility of each text since there is one column provided for both categories.
R2	**On the result of the translation:** On T1, the rater acknowledge some missing aspects such as articles and particles. The rater also believes that too many simple sentences are used in the this text. In summary, T1 text is considered quite readable. On T2, the rater gives positive feedback with the overall score of 3. **On the rubric:** The rater feels that the rubric is not following a common framework. She

*Author(s) Correspondence:
E-mail: ataufik@bundamulia.ac.id

Versi Online: http://journal.ubm.ac.id/xxx/xxx
DOI : dx.doi.org/xxx/xxxxxxxx
Hasil Penelitian

Journal of English Language and Culture
Vol. 10 (No. 2) : 64 - 74. Th. 2020
p-ISSN: 2087-8346
e-ISSN: 2597-8896

	believes it needs to be further modified. In addition, she notices that there are no instructions provided for the rubric.
R3	**On the result of the translation:** On T1, the rater believes it to be not a good translation; it is considered not natural. She also says that the structure of the compound and complex sentences, and lack of punctuation, contribute to the confusion. She further adds that the translation is quite understandable until a certain point. On T2, she mentions the visibility of ST in the word 'where'. This highlighted word is one of the revision made in the pre-editing process. **On the rubric:** The rater shares other rater's opinion on the unfriendly nature of the rubric; it is believed to be formatically confusing.
R4	**On the result of the translation:** The rater mentions that T2 is much better than T1. He, however, adds that in overall, both results are still confusing; the flow of the writing is not present in both. **On the rubric:** -

From the early findings, it can pre-concluded that the pre-edited version is slightly better in terms of naturalness and readability with a difference in point of 0.875. One major finding that is consequential to the hypotheses is on the addition of conjunction. After being added with conjunction, it is found that certain conjunction does not translate well in MT, such as the conjunction 'dimana'. This conjunction may have the subordinative meaning in the ST, however, in the TT, it is translated literally into 'where', which is not an appropriate conjunction for the sentences in the TT. Another interesting finding from the Human Evaluation is that the flow of writing in the MTs is not present. This shows that the MT is still unable to create a good flow of translation as compared to human translation.

With such findings, an additional hypothesis can be added.

H3
In the Pre-editing process, avoid using conjunctions which are ST specific, try to use generally accepted conjunctions.

In addition, the negative comments are actually on the evaluation model itself. Based on the feedbacks from the human evaluation, the rubric has been revised into the following:

Table 4. Revised Modified Three-Grade Scale of Translation Evaluation

Score	Text	Score Input Put a check (√)	Naturalness	Readibility
3	T1		It is very natural. Structural naturalness is achieved.	It is very easy to read
	T2			
2	T1		It is acceptable. Source Language structure is still visible.	.It is quite difficult to read. Much efforts are needed.
	T2			
1	T1		It is not a good translation.	It is not readable
	T2			

Instruction: Put a check (√) in the corresponding column based on the text provided.

*Author(s) Correspondence:
E-mail: ataufik@bundamulia.ac.id

Versi Online: http://journal.ubm.ac.id/xxx/xxx
DOI : dx.doi.org/xxx/xxxxxxxx
Hasil Penelitian

Journal of English Language and Culture
Vol. 10 (No. 2) : 64 - 74. Th. 2020
p-ISSN: 2087-8346
e-ISSN: 2597-8896

Having revised the evaluation model, the researcher tests the hypotheses on other text. The following is the ST:

ST

Idiom identik dengan kesukaran pemaknaan melalui makna harfiahnya. Kesulitan ini pada dasarnya muncul karena adanya konstituen dalam idiom yang tidak menyumbangkan maknanya untuk makna idiom secara keseluruhan. Secara umum, jenis idiom berdasarkan makna dapat terbagi menjadi tiga, yaitu idiom komposisional, komposisional sebagian, dan nonkomposisional. Idiom komposisional sebagian dan nonkomposisional bersifat metaforis. Idiom metaforis seringkali diterjemahkan tanpa memaknai metafora yang dikandungnya, seperti yang dinyatakan oleh Knowles dan Moon (2005:16) "In normal contexts, we are likely to interpret their idiomatic meanings without thinking about the metaphors that they contain" dengan contoh "The monthly payments cost ***an arm and a leg****" yang berarti 'Pembayaran bulanan* ***sangat mahal****.' Padahal, penggunaan idiom metaforis berfungsi untuk menyampaikan pesan tertentu dari pembicara kepada kawan bicaranya. Pesan tersebut tidak dapat diterka jika hanya memaknai arti harfiah dari idiomnya saja, tetapi juga dengan memahami metafora yang dikandungnya, serta konteks pemakaian metafora tersebut. Penelitian ini berusaha menentukan karakteristik semantis idiom metaforis, pemakaian idiom, serta nilai budaya dan sosial yang tercermin di dalam konteks. Penelitian ini diharapkan dapat memperkaya khasanah penelitian metafora dalam bahasa Mandarin. Berbeda dengan penelitian sebelumnya yang menemukan pola idiom, struktur idiom, penelitian ini akan memberikan sumbangan yang agak berbeda yaitu konteks pemakaian, sehingga penelitian ini dapat memberikan sumbangan mengajarkan idiom dalam pengajaran bahasa Mandarin. Agar data idiom yang digunakan dalam penelitian ini mewakili idiom yang masih populer, maka digunakan idiom-idiom empat-karakter yang terdapat dalam korpus daring "Chinese Internet Corpus". Data yang diperoleh dari korpus tersebut akan dicek kembali untuk memastikan bentuk empat-karakter tersebut adalah idiom bahasa Mandarin. Kemudian akan dilakukan pemetaan idiom berdasarkan ranah sumber (source domain) dan ranah sasaran (target domain). Penelitian yang bersifat deksriptif kualitatif ini akan mengungkapkan konteks pemakaian idiom berdasarkan bentuk idiom yang mengandung kosakata bilangan dan tubuh manusia.*

And the Raw TT:

Idiom is identical with difficulty of meaning through its literal meaning. This difficulty basically arises because there are constituents in idioms that do not contribute their meaning to the meaning of the idiom as a whole. In general, types of idioms based on meaning can be divided into three, namely compositional, partial compositional, and noncompositionional idioms. Partial and non-compositional idioms are metaphorical. Metaphorical idioms are often translated without interpreting the metaphors they contain, as stated by Knowles and Moon (2005: 16) "In normal contexts, we are likely to interpret their idiomatic meanings without thinking about the metaphors that they contain" with the example "The monthly payments cost an arm and a leg "which means 'monthly payments are very expensive.' In fact, the use of metaphorical idioms serves to convey certain messages from the speaker to his interlocutors. The message cannot be guessed if it only interprets the literal meaning of the idiom, but also by understanding the metaphors it contains, and the context of the use of the metaphor. This study seeks to determine the semantic characteristics of metaphorical idioms, the use of idioms, as well as cultural and social values that are reflected in context. This research is expected to enrich the realm of metaphor research in Mandarin. In contrast to previous studies that found idiom patterns, idiom structures, this study would make a somewhat different contribution

*Author(s) Correspondence:
E-mail: ataufik@bundamulia.ac.id

Versi Online: http://journal.ubm.ac.id/xxx/xxx
DOI : dx.doi.org/xxx/xxxxxxxx
Hasil Penelitian

Journal of English Language and Culture
Vol. 10 (No. 2) : 64 - 74. Th. 2020
p-ISSN: 2087-8346
e-ISSN: 2597-8896

to the usage context, so that this research could contribute to teaching idioms in Mandarin teaching. In order for the idiom data used in this study to represent popular idioms, the four-character idioms used in the Chinese Internet Corpus online corpus are used. Data obtained from the corpus will be checked again to ensure the four-character form is a Chinese idiom. Then idiom mapping will be based on the source domain and target domain. This descriptive qualitative research will reveal the context of the use of idioms based on idiom forms that contain vocabularies of numbers and the human body.

As it is with bilingual analysis, on the comparison between the ST and Raw MT, one of the confusing parts (mistranslation) is on the sentence *The message cannot be guessed if it only interprets the literal meaning of the idiom, but also by understanding the metaphors it contains, and the context of the use of the metaphor*. Which is an MT result of *Pesan tersebut tidak dapat diterka **jika** hanya memaknai arti harfiah dari idiomnya saja, tetapi juga dengan memahami metafora yang dikandungnya, serta konteks pemakaian metafora tersebut*. According to the researcher, this is caused by the inappropriate use of Indonesian correlative conjunction of 'tidak hanya ... tetapi (juga)...', which is similar to the correlative conjunction of 'not only ... but also ...'. As can be seen from the highlighted part, the ST author is using 'jika' instead of 'dari' which is more appropriate. From this, another hypothesis can be added.

H4
If we use the appropriate correlative conjunction, the result of MT will be more appropriate.

The second mistake is seen from the use of the word 'menemukan' which is translated into 'found' in the Raw MT.

ST
*Berbeda dengan penelitian sebelumnya yang **menemukan** pola idiom, struktur idiom, penelitian ini akan memberikan sumbangan yang agak berbeda yaitu konteks pemakaian, sehingga penelitian ini dapat memberikan sumbangan mengajarkan idiom dalam pengajaran bahasa Mandarin*
ST
*In contrast to previous studies that **found** idiom patterns, idiom structures, this study would make a somewhat different contribution to the usage context, so that this research could contribute to teaching idioms in Mandarin teaching*

As seen from the highlighted parts, the word 'menemukan' is better if it is changed into 'berfokus'. As it is, another hypothesis can be added.

H5
Pre-editing process must also focus on inappropriate wordings.

Having found new hypotheses, the researcher is unable to test the previous hypotheses since there are (1) overly long sentences, and (2) inappropriate or lack of conjunctions. Therefore, the researcher moves on to the next process, which is to create the PrE ST, and the following is the result of the process.

PrE ST
*Idiom identik dengan kesukaran pemaknaan melalui makna harfiahnya. Kesulitan ini pada dasarnya muncul karena adanya konstituen dalam idiom yang tidak menyumbangkan maknanya untuk makna idiom secara keseluruhan. Secara umum, jenis idiom berdasarkan makna dapat terbagi menjadi tiga, yaitu idiom komposisional, komposisional sebagian, dan nonkomposisional. Idiom komposisional sebagian dan nonkomposisional bersifat metaforis. Idiom metaforis seringkali diterjemahkan tanpa memaknai metafora yang dikandungnya, seperti yang dinyatakan oleh Knowles dan Moon (2005:16) "In normal contexts, we are likely to interpret their idiomatic meanings without thinking about the metaphors that they contain" dengan contoh "The monthly payments cost **an arm and a leg**" yang berarti 'Pembayaran bulanan **sangat mahal**.'*

*Author(s) Correspondence:
E-mail: ataufik@bundamulia.ac.id

Versi Online: http://journal.ubm.ac.id/xxx/xxx
DOI : dx.doi.org/xxx/xxxxxxxx
Hasil Penelitian

Journal of English Language and Culture
Vol. 10 (No. 2) : 64 - 74. Th. 2020
p-ISSN: 2087-8346
e-ISSN: 2597-8896

Padahal, penggunaan idiom metaforis berfungsi untuk menyampaikan pesan tertentu dari pembicara kepada kawan bicaranya. Pesan tersebut tidak dapat diterka hanya ***dari*** *memaknai arti harfiah dari idiomnya saja, tetapi juga dengan memahami metafora yang dikandungnya, serta konteks pemakaian metafora tersebut. Penelitian ini berusaha menentukan karakteristik semantis idiom metaforis, pemakaian idiom, serta nilai budaya dan sosial yang tercermin di dalam konteks. Penelitian ini diharapkan dapat memperkaya khasanah penelitian metafora dalam bahasa Mandarin. Berbeda dengan penelitian sebelumnya yang* ***berfokus pada*** *pola idiom, struktur idiom, penelitian ini akan memberikan sumbangan yang agak berbeda yaitu konteks pemakaian, sehingga penelitian ini dapat memberikan sumbangan mengajarkan idiom dalam pengajaran bahasa Mandarin. Agar data idiom yang digunakan dalam penelitian ini mewakili idiom yang masih populer, maka digunakan idiom-idiom empat-karakter yang terdapat dalam korpus daring "Chinese Internet Corpus". Data yang diperoleh dari korpus tersebut akan dicek kembali untuk memastikan bentuk empat-karakter tersebut adalah idiom bahasa Mandarin. Kemudian akan dilakukan pemetaan idiom berdasarkan ranah sumber (source domain) dan ranah sasaran (target domain). Penelitian yang bersifat deksriptif kualitatif ini akan mengungkapkan konteks pemakaian idiom berdasarkan bentuk idiom yang mengandung kosakata bilangan dan tubuh manusia.*

And the PrE TT:

PreTT
Idiom is identical with difficulty of meaning through its literal meaning. This difficulty basically arises because there are constituents in idioms that do not contribute their meaning to the meaning of the idiom as a whole. In general, types of idioms based on meaning can be divided into three, namely compositional, partial compositional, and noncompositionional idioms. Partial and non-compositional idioms are metaphorical. Metaphorical idioms are often translated without interpreting the metaphors they contain, as stated by Knowles and Moon (2005: 16) "In normal contexts, we are likely to interpret their idiomatic meanings without thinking about the metaphors that they contain" with the example "The monthly payments cost an arm and a leg "which means 'monthly payments are very expensive.' In fact, the use of metaphorical idioms serves to convey certain messages from the speaker to his interlocutors. The message cannot be guessed only from interpreting the literal meaning of the idiom, but also by understanding the metaphors it contains, and the context of the use of the metaphor. This study seeks to determine the semantic characteristics of metaphorical idioms, the use of idioms, as well as cultural and social values that are reflected in context. This research is expected to enrich the realm of metaphor research in Mandarin. In contrast to previous studies that have focused on idiom patterns, idiom structures, this study will make a somewhat different contribution to the usage context, so that this research can contribute to teaching idioms in Mandarin teaching. In order for the idiom data used in this study to represent popular idioms, the four-character idioms used in the Chinese Internet Corpus online corpus are used. Data obtained from the corpus will be checked again to ensure the four-character form is a Chinese idiom. Then idiom mapping will be based on the source domain and target domain. This descriptive qualitative research will reveal the context of the use of idioms based on idiom forms that contain vocabularies of numbers and the human body.

Having gained the PrE TT, using the same rater, the Human Evaluation is conducted. This time, based on the feedback on the first testing, the focus area is highlighted. The followings are the results:

*Author(s) Correspondence:
E-mail: ataufik@bundamulia.ac.id

Versi Online: http://journal.ubm.ac.id/xxx/xxx
DOI : dx.doi.org/xxx/xxxxxxxx
Hasil Penelitian

Journal of English Language and Culture
Vol. 10 (No. 2) : 64 - 74. Th. 2020
p-ISSN: 2087-8346
e-ISSN: 2597-8896

Table 5. Result of Human Evaluation for Academic Text of Raw TT and PrE TT

Raters/Texts	T1 (on Naturalness and Readability)	T2 (on Naturalness and Readability)
R1	2	3 and 2
R2	2	3
R3	2 and 2	3 and 2
R4	2 and 3	3
Average	**2.125**	**2.75**

From the results, it can be seen that the PrE MT gained better result of evaluation with a difference of point of 0,625. In addition, there is also ffedback given by the raters (R1 and R3) on the Model of Evaluation. Both suggested that the description of Naturalness and Readability is differentiated into the followings.

Table 6. Second Revision of Modified Three-Grade Scale of Translation Evaluation

	T1 Score	T2 Score
Naturalness		
Readability		

Score	Description of Naturalness	Description of Readability
3	It is very natural. Structural naturalness is achieved.	It is very easy to read.
2	It is acceptable. Source Language structure is still visible.	.It is quite difficult to read. Much efforts are needed.
1	It is not a good translation	It is not readable.

CONCLUSION(S)

From the testing of pre-editing of the MT and Human Evaluation, several interesting conclusions can be made. The first is related to the rules. As seen from the number of hypotheses in each testing, these are some of the rules of pre-editing:

1. Do not bother with field-specific terminologies since a deeper level of research on it must be implemented.
2. Divide longer sentences into a more comprehensible units
3. If possible, on complex sentences clarify the sub-ordinating conjunctions, and only use generally acceptable (not specific) ones.
4. Pay attention to correlative conjunctions
5. Pay attention to unappropriate words in ST and change them to an appropriate ones.

SUGGESTIONS

On the Human Evaluation model, the very simple model suggested by Hiraoka and Yamada (2019) is simply not enough. Even after being edited with a somewhat detailed descriptions, the raters are still having difficulty with the evaluation. In addition, in the researcher's opinion, there is a need to rate the validity and reliability of the Second Revision of Modified Three-Grade Scale of Translation Evaluation to achieve a more

*Author(s) Correspondence:
E-mail: ataufik@bundamulia.ac.id

Versi Online: http://journal.ubm.ac.id/xxx/xxx
DOI : dx.doi.org/xxx/xxxxxxxx
Hasil Penelitian

Journal of English Language and Culture
Vol. 10 (No. 2) : 64 - 74. Th. 2020
p-ISSN: 2087-8346
e-ISSN: 2597-8896

reliable and valid results of the testing of the process.

On the rules themselves, a further research needs to be conducted to further develop the research. This is evidence in the rise of new hypotheses on each of the testing. In addition, although the results of Raw TT and PrE TT are not that significantly different, the fact is still true; the PrE is, overally, better, at least in naturalness and readability.

REFERENCES

Gerlach, J., et al. (2013). Combining pre-editing and post-editing to improve SMT of user-generated content. In: *O'Brien, S., Simard, M. & Specia, L. Proceedings of MT Summit XIV Workshop on Post-editing Technology and Practice*. p. 45-53.

Hiraoka, Y., & Yamada, M. (2019). Pre-editing plus neural machine translation for subtitling: Effective pre-editing rules for subtitling of TED Talks. *MT Summit* (2): 64-72.

Miyata, R., & Fujita, A. (2017, May). Dissecting human pre-editing toward better use of off-the-shelf machine translation Systems. In *Proceedings of the 20th Annual Conference of the European Association for Machine Translation (EAMT)*, User studies papers.

Shei, C.-C. (2002) Teaching MT through pre-editing: Three case studies. In: *6th EAMT Workshop Teaching Machine Translation*. Manchester, 89-98.

Statt, N. (2016, September). Google's AI translation system is approaching human-level accuracy: But there's still significant work to be done. *The Verge*. Retrieved from https://www.theverge.com/2016/9/27/13078138/google-translate-ai-machine-learning-gnmt

*Author(s) Correspondence:
E-mail: ataufik@bundamulia.ac.id