Machine Translation and Fixed Expressions

Alvin Taufik
Universitas Bunda Mulia
ataufik@bundamulia.ac.id

Abstract

The objective of this research is to find out the potentiality of machine based translation, such as *Google Translation* for example, in translating fixed expressions. The framework for fixed expression will be based on Baker's typology of equivalence (1992) and Alexander's classification (1978). The research will then compare the result of the translation done by the machine and the ones made by human translators which is readily available in the public. The Source Language is in Bahasa Indonesia while the Target Language will be in English. The result of the research shows that Google has the possibility to translate morphological and syntactical level units of language, yet has no to little possibility to translate Sentence level unit of language.

Keywords: *Machine Translation, Fixed Expressions*

1. Introduction

1.1. Background

How does machine translation work? To answer the question, it is best if we know what is meant first by Machine Translation (MT). To differentiate between MT and Computer Aided Translation, Gonzales (2012) has written an interesting separation between them. She mentioned that a human factor in one differentiate the almost non existence in the other. In Computer Aided Translation (CAT) all the hard work is still done by the HUMAN translator with the help of translation memory (TM). This memory is used to optimize consistency of the terminology used by translator/s in their translation. In Machine Translation (MT) human's translation results are used as the basis of its often statistical and rule-based approach. Google Translate software is one example of the latter.

So how does Google translate work? In the 'About' page of Google Translate software, this is written:

"When Google Translate generates a translation, it looks for patterns in hundreds of millions of documents to help decide on the best translation for you. By detecting patterns in documents that have already been translated by human

translators, Google Translate can make intelligent guesses as to what an appropriate translation should be. This process of seeking patterns in large amounts of text is called "statistical machine translation" (2014).

The translations, then, are not perfect since the machine is doing all the translation based on the number of available data that the software can refer to; in other word, the more human-translated documents become the reference, the better the translation will be.

As it relies on statistical choices, the data corpus is very limited if compared to units of language in general. Language itself comprised of smaller units, such as words, phrases, sentences, clauses, and others. In this research the focus of the experiments will be on fixed expressions. Fixed expressions are chosen as the subject for its rigidness. Unlike other units of language, their pattern is 'frozen' (Baker, 1992). Even though they have 'frozen' pattern, fixed expression also have fairly transparent meaning. However, the meaning of fixed expressions cannot be concluded from the collection of meaning of each word; they must be taken as a unit of meaning. It is for this reason that the researcher is interested in finding out whether machined-translation has the capability to translate fixed expression as a unit

As it has been mentioned previously, with the vast amount of resources which become the basis of the choice for the machined-translation, it is still small in comparison with the units of language available and in use by people, whether it is in their conversation, or in their writings. In the case of fixed expressions, although they are fairly literal in meaning, the component of meaning itself is taken from the whole expressions as a unit. It is interesting whether or not the machine translation would be able to properly translate the fixed-expressions, given the fact. As it is, it can be formulated the following questions:

1. How accurate and natural is the machine translation in translating fixed expressions?
2. What can be done from the results of the translation by the machine?

The research is aimed to find out whether or not the machine translation has the possibility to appropriately translate fixed expression. It is significant since, if the machine is able to properly translate the fixed expressions, it can be used in

transferring any fixed expressions in Indonesian into English at anytime. However, if the machine is incapable of transferring the fixed expressions properly, this research can identify the problem area of the translation done by the machine.

2. Literature Review

2.1. Fixed Expressions

Baker (1992) stated that fixed expressions are 'extreme' collocations. It is extreme since they allow only a little or no variations. Unlike collocation which has various coupling, as the example of the word 'homework' which can collocate with the verbs 'do', 'have' and others, such is not the case with fixed expressions. Furthermore, Baker also states that fixed expression is a competence which cannot be matched by foreign speakers of the language; in other words, it is native speakers' trait to use more fixed expressions.

In translation, the main problem of fixed expressions is in the ability of recognizing and interpreting them, and in conveying the various aspects of meaning in the expressions itself. After the recognition, in translating fixed expressions, there are problems which could hinder the process of rendering. One of the difficulties is, they may have no equivalence in the TL. Let's have a look at two examples of fixed expressions:

A1. Ladies and Gentlemen

B1. Further to your letter of…..

For the first expression, in the Target Language, Namely Bahasa Indonesia, there are appropriate units of equivalence, that is:

A2. Bapak dan Ibu sekalian, Bapak dan Ibu yang saya hormati, etc

However, in the latter expression, it cannot be translated per unit as in the first. Instead they need to be taken as a unit. Even then, there is a possibility of it not having the proper equivalence. Thus, it can be translated as a whole into:

B2. Menjawab surat bapak/ibu mengenai….

The second problem of translating fixed expressions is the different context in which they are used. Unfortunately, the researcher cannot provide explanations on this.

What Baker did not mention specifically in her book, is a detailed description of fixed expressions themselves. Alexander (1978) pointed out several categories of fixed expressions, they are:

1. Idioms, which are further classified into idioms made of phrasal verbs, 'tournures', and irreversible binomials,
2. Discourse-structuring devices, such as the ones used in greetings (long time no see) or connectives/gambits (for a kick off),
3. Proverbs,
4. Catchphrases, such as clichés and slogans, and
5. Quotations and allusions

Since the research is focused on fixed expressions which exist in both the Source Language in Bahasa Indonesia, and its Target Language in English, only two of the five categories of fixed expressions as stated by Alexander will be discussed in this paper. They are Idioms and Proverbs.

In addition, Lusiana (2012) has also made some classifications on idioms in Bahasa Indonesia. The first category of idioms according to her is Idioms related with body parts. The second type is idioms which using human senses in its expressions. The third type of idiom is the ones which are related to nature. The fourth category is idioms which are related to color. The fifth type is related to animals and plants, while the last are idioms which involve numbers.

For the purpose of accessibility and clarity, the second classification, discourse structuring devices will not be discussed in this research since there is no clear definition of it in Alexander's explanation. In his third classification, Alexander mentions proverbs. Encyclopedia Britannica stated that a proverb is 'succinct and pithy saying in general use, expressing commonly held ideas and beliefs'. Furthermore, in Indonesia, a more comprehensive dictionary on proverbs has been made by Badudu (2008).

2.2. Machined-Translation (Google Translate)

As software originally created in the 1980's, it is specifically designed to perform intellectual presupposition of an expression from its syntax and vocabulary (Bellos, 2011). Google Translate (GT) can scan a corpus of data which range all the way to the paper produced since 1957 by the EU in twenty four different languages, and a lot more such as the records of international tribunals to company reports and all the articles and books in bilingual form that have been uploaded on the web by individuals, libraries, booksellers, authors and academic departments. Moreover, he also mentioned that the way GT works emphasize on the way of thinking that the language the human used are basically similar and they have been reused over and over again.

3. Methodology

The research is a qualitative research since the purpose of the research is basically to provide description of whether or not Google Translate is able to provide proper Target Language (TL) for fixed expressions' translation. The research is comparative in nature since it will compare the result of the translation made by the machine with the ones made by human translators which have been readily available in different types of texts which contains the fixed expressions as specified in this research.

The data will be gathered by inputting different types of fixed expression as stated by Alexander (1978) and compared the results of the translation done by the machine and the actual corpus available in the public.

Regarding the samples, since there is a really vast amount of usage of fixed expressions in the public, the sampling procedure will be purposive sampling. The researches will choose a Source Language (SL) data which has its TL counterpart, and match them with the specification made by Alexander concerning fixed expressions.

4. Data Analysis

4.1. Idioms

In Bahasa Indonesia, idiom is defined as a combination of words with meanings unlike the original words meanings. Lusiana (2012) classified idioms in Bahasa Indonesia into several types, namely the ones using body parts, using human senses, the ones related to colors, nature, numbers, plants and animals.

4.1.1. Idioms with body parts

Here are some examples of idioms using body parts:

a. Jeng Sri memang *tinggi hati*.(sombong)
b. Karena ucapan orang itu, Waluyo *naik darah*.(marah)
c. Itulah akibatnya kalau menjadi anak yang *berkepala batu*. (tidak mau menurut)

In all examples, the ones in italics are the idioms. In the first example above, the meaning of *tinggi hati* is arrogant. In the second example, the meaning of *naik darah* is to become angry. The meaning of the third idiom in the example is stubborn. As can be seen from the context, they are attributed to a person's characteristics. Here are the results of the translation made by Google Translate, and its comparison of naturalness as seen from COCA.

Idioms	SL context	TL context (from COCA)
Tinggi hati	Jeng Sri memang *tinggi hati*	My own professors were fairly *haughty* and aloof
Naik darah	Karena ucapan orang itu, Waluyo *naik darah*	And that made the angry mob *seered*
Kepala batu	Itulah akibatnya kalau menjadi anak yang *berkepala batu*.	He fixed me with that luminous gray-eyed stare that conveyed *implacable* determination

Table 1. Translation of Idioms using Body Parts

As can be seen from the comparison, the idioms as combination of words are not translated as such in the TL. From the example only one idiom is possible to be translated into another idiom. From this, it can be concluded that idioms in Bahasa Indonesia can and will be translated into its matching idioms in English if they are commonly used. From the context it can be seen that the appropriate

meaning is also transferred successfully, although in one of the example, there is a slight change in the sense of the meaning. In Indonesia, the idiom *kepala batu* has a negative meaning. Meanwhile, in the TL, as seen from other examples in COCA, it has a positive meaning which is often related with determination.

4.1.2. Idioms using Human Senses

The followings are some examples of Idioms with Human Senses.

a. Hati-hati terhadap orang yang *besar mulut* itu. (suka membual)
b. *Merah telinganya* ketika ia dituduh sebagai koruptor. (marah)
c. Karena *gelap mata*, dia mengamuk di kantor. (hilang kesabaran)

In the italicized idioms, the meaning of *besar mulut* is similar to boastful, while *Merah telinga* has the meaning of getting angry. The third idioms in the example, *gelap mata* has the meaning of losing patience. These are the results of the translation:

Idioms	SL context	TL context (from COCA)
Besar mulut	Hati-hati terhadap orang yang *besar mulut* itu	I was *vainglorious* and thought not of myself as a man who could be hurt and could die
Merah telinga	*Merah telinganya* ketika ia dituduh sebagai koruptor	*Red ear* (no contextual similarity)
Gelap mata	Karena *gelap mata*, dia mengamuk di kantor	*Dark eyes* (no contextual similarity)

Table 2. Translation of Idioms using Senses

In this type of idiom, notice that two out of three samples presented here was translated very literally. Since they are translated literally, it is obvious that their uses are not common in the TL. This is proven after the results of the translation or the TL is compared to COCA database. Since *Google Translate*is based primarily on the search engine database, it can be concluded that the database on idioms in Bahasa Indonesia which are related to human senses is not as extensive as the other types of idiom. As additional information, the common idiom used to describe anger which is related to ears in English is *burning ear*.

4.1.3. Idioms related with Colors

The words stated below are some examples of idioms related to colors.

a. Lebih baik *berputih tulang* daripada hidup menanggung malu seperti ini. (mati)
b. Ketika kutinggalkan dulu engkau *masih merah*, sekarang sudah seorang jejaka. (masih bayi)

Here are some comparison of the SL and TL:

Idioms	SL context	TL context (from COCA)
Putih tulang	Lebih baik *berputih tulang* daripada hidup menanggung malu seperti ini	*White bone* (no contextual similarity)
Masih merah	Ketika kutinggalkan dulu engkau *masih merah*, sekarang sudah seorang jejaka	*stillred*(no contextual similarity)

Table 3. Translation of Idioms related with Colors

As with the previous type, *Google Translate* has translated the idioms word per word. Also in line with previous findings, the result of the translation is not commonly used in the TL context. It can be concluded that idioms related with colors cannot be translated appropriately.

4.1.4. Idioms related with Nature

What is meant by nature here are things which can be found in nature. The words stated below are some examples of idioms related to nature.

a. Selama pertandingan sepak bola itu, benar-benar dia menjadi *bintang lapangan*. (pemain yang baik)

From the sample above *bintang lapangan* is the idiom related with nature. The meaning of such idiom is about the best (star) player in the field. Here is the translated version of the idiom.

Idioms	SL context	TL context (from COCA)
Bintang lapangan	Selama pertandingan sepak bola itu, benar-benar dia menjadi *bintang lapangan*	Thalia is coursing through a relatively sparse *starfield* in the constellation Virgo

Table 4. Translation of Idioms related to Nature

Seen from the result, although there is an actual context for the TL, its meaning is world apart. The translation also reveals that the result of Google Translate shows a different word group. What in the SL is an idiom has been changed into a simple noun phrase in the TL. In terms of meaning, in the SL the idiom has the meaning of 'best player'; the matching phrase for that in the target language is 'star athlete'. The result of the translation, however, is related to galaxy or galactic beings. Star field can be loosely defined as a gathering of stars, which exist in a galaxy. It can be concluded then that idioms with things related to nature cannot be naturally translated into English.

4.1.5. Idioms related with Animals and Plants.

These types of idioms are the ones which incorporated animals and plants in its expressions. Here are some examples of them.

a. Lagi-lagi aku yang *dikambing hitamkan* bila timbul keributan di kelas
b. "Gema Tanah Air" sebuah *bunga rampai* yang disusun oleh H.B. Jassin.

And, here are the results of the translations:

Idioms	SL context	TL context (from COCA)
Kambing hitam	Lagi-lagi aku yang *dikambing hitamkan* bila timbul keributan di kelas	when something goes wrong, like they always do, they will have a convenient *scapegoat*
Bunga rampai	"Gema Tanah Air" sebuah *bunga rampai* yang disusun oleh H.B. Jassin.	Historical Quarterly completed its 94th year of publication, offering readers in 2010 a veritable*potpourri* of Georgia history

Table 5. Translation of Idioms related to Plants and Animals

As seen from the results above the idioms in Bahasa Indonesia has been translated successfully into English. Although the results are not an equal word class of the counterparts, that is not the issues, since the focus is on whether Google Translate has the ability to translate such fixed expressions. The other focus is also on the naturalness of the translation. According to the context seen in the TL, it can be seen that naturalness has been achieved in the translation. Thus it

can be concluded that idioms with animals and plants in its expression can be transferred naturally.

4.1.6. Idioms related with Numbers

Here are some examples of the last type of idioms according to Lusiana; idioms related with numerical values.

a. Kalau bekerja dengan *setengah hati*, hasilnya kurang memuaskan.

And, here is the result of the translation.

Idioms	SL context	TL context (from COCA)
Setengah hati	Kalau bekerja dengan *setengah hati*, hasilnya kurang memuaskan	She turned away with the air of someone *halfheartedly* resigned to endure, but as she turned, she started and stopped

Table 6. Translation of Idioms related to Plants and Animals

From the results above, it can be seen that there is a common expression in the TL which uses similar numerical expressions. In addition, if seen from the context, the meaning is similar between the SL and the TL. It can be concluded then that Google Translate has the potential to translate this type of idioms accurately and naturally.

4.2. Proverbs

In his third classification, Alexander mentions that proverb is also one type of fixed expressions. Badudu (2008) has created a dictionary on proverbs. For the practicality of this research, five items will be chosen to become the first samples. If from the five the results have been identified, there will be no choosing of other samples. However, if the first five samples show inconclusive results, further sampling will be conducted.

Proverbs (SL)	**TL (Google Translate)**
Kalah jadi abu, menang jadi arang	lostto ashes, winsocharcoal
Bermain air basah, bermain api hangus	waterplaywet, playing with firescorched
Air tuba dibalas dengan air susu	Watertuberewardedwithmilk

Lempar batu sembunyi tangan	Throwrocksto hide the hand
Berat sama dipikul, ringan sama dijinjing	Bear the same weight, same lightweight portable

Table 7. Translation of Proverbs

From the result, it can be seen clearly that even without checking with COCA, the TL version is erroneous in both equivalent and grammatically. Thus it can be concluded that proverbs translation cannot have equivalent translation if Google Translate is used.

5. Conclusion

As can be seen from the findings in chapter five, there are several points which can be concluded. The first is on the accuracy and naturalness of Google Translate in rendering fixed expressions in Indonesia into English. This can be categorized into two parts; first as it deals with idiomatic expressions and second on proverbs.

From all the findings on idiom translation, it can be concluded that not all idioms can be translated properly. In this research, especially on idioms related to nature, colors and senses, Google Translate does not have the potential to render it accurately and naturally yet. It can also be concluded that the database on idioms related to those specifics mentioned are still lacking in the search engine.

In relation to Proverbs, it can be concluded that Google Translate still has no potential to translate sentence level unit of language. GT still translate the proverbs literally and the error in grammar is still a factor in the Target Language. Concerning the second research question, on what can be done to the TL results of Google Translate, what can be done is to input more feedback on Google Translate concerning those two types of expressions. This feature is actually readily available currently in Google Translate itself.

References

Alexander, R. J. Fixed expressions in English: Reference books and the teacher, *English Language Teaching Journal*, *38*(2), 127-134.

Badudu, J. S. (2008). *Kamus peribahasa*. Jakarta: Penerbit Buku Kompas.

Baker, M. (1992). *In other words: A coursebook on translation*. London: Routledge.

Bellos, D. (2014, October 7). How Google Translate works. *The Independent*.

Gonzales, A. Computer aided translation vs machine translation (TM vs MT). [Web log]. Accessed on August 16, 2012.

Lusiana, I. Ungkapan/Idiom in *Materi Bahasa Indonesia*. Accessed on January 19, 2015.