

# YOUTUBE'S AUTOMATED TRANSLATION AND GOOGLE TRANSLATE: POSSIBLE PROBLEMS AND SOLUTIONS

Alvin Taufik\*

English Language and Culture Department, Bunda Mulia University

Received on 21 September 2018 / Approved on 15 October 2018

## ABSTRACT

Translation technology has now been acknowledged as a subdomain of translation studies (Christensen, Flanagan, and Schjoldager, 2017). The researches on translation technology cover a lot of aspects of translation as well as the technology that is used to aid the process. This research is focused on the latter, especially on the results of the Human-aided Machine Translation (HMT). The HMT which becomes the focus of this research are Youtube's Automated Translation and Google Translate (AT and GT respectively). The aim of this research is to evaluate the HMTs, identify the problems, and propose a possible solution for the problems. This research is using a realism approach because it uses the translator students' perspective of the HMT, and the use of corpora. Questionnaires and experiments will be given and performed to assess and to formulate the solutions.

**Keywords:** HMT, GT, AT, realism, corpora

## ABSTRAK

Teknologi penerjemahan kini telah diakui sebagai subdomain dari studi terjemahan (Christensen, Flanagan dan Schjoldager, 2017). Penelitian tentang teknologi terjemahan mencakup banyak aspek penerjemahan serta teknologi yang digunakan untuk membantu proses penerjemahan. Penelitian ini difokuskan terutama pada hasil Terjemahan Mesin dengan Bantuan Manusia (Human-Aided Machine Translation/HMT). HMT yang menjadi fokus penelitian ini adalah Terjemahan Otomatis YouTube dan Google Translate (AT dan GT). Tujuan dari penelitian ini adalah untuk mengevaluasi HMT, mengidentifikasi masalah, dan mengusulkan solusi untuk masalah tersebut. Penelitian ini menggunakan pendekatan realisme karena menggunakan perspektif mahasiswa penerjemah tentang HMT, dan penggunaan korpora. Kuesioner dan eksperimen akan diberikan dan dilakukan untuk menilai dan merumuskan solusi.

**Kata Kunci:** HMT, GT, AT, realisme, korpora

## INTRODUCTION

### 1.1. Background

As our daily life demands more translational activity, so does the translator's desire to keep up with the demands. For many years, technology has been used in assisting many aspects of the people's life. The world of translation has also embraced the technology to help with its practice. One of the most dominant and frequently used of translation technology is the human-aided machine

translation (HMT). Two of the many HMT technology, and also the most well known, is Youtube's Automatic Translation (AT) and Google Translate (GT). As the title suggest, this research aims to find out the problems in Youtube's AT and GT, and investigate the possible solutions to deal with the problems.

The reasons behind the choosing of both sites are because they are readily available, free, user-friendly, and fast. What is meant by fast is that both system can render one source language (SL) into target language

\*Author(s) Correspondence:

E-mail: [ataufik@bundamulia.ac.id](mailto:ataufik@bundamulia.ac.id)

(TL) really fast. This is obviously beneficial for translation practitioner.

The other reasons why the two sites are chosen is because of the two sites' connectivity. It is a known fact that Youtube has been bought by Google, so it is interesting to find out whether the purchase has any effects on the system's functionality.

On a further note, Youtube's Automated Translation (AT) and Google Translate (GT) have their own purposes. While GT is occasionally used to translate written text, Youtube's AT is used to translate audiovisual materials. This distinction makes a significance difference in the way the two ATs are being evaluated. While text and its embedded values become the foci of GT, Youtube's AT has to consider more factors such the image shown on the screen, and often the accompanying sounds. With those factors in mind, it is interesting to see which system can produce the most acceptable results in their respective domains. In addition, the language pair which will be the focus of this research is the language pair of Indonesian and English.

## 1.2. Statement of Problems and Research Question/s

To achieve the desired results from the research, several problems need to be identified and addressed. First, the researcher needs to find most suitable situation to evaluate the results fairly and accordingly. Only when the problems have been dealt accordingly, the research aims to identify the problems which exist in the translation results of the two systems, and to propose possible solutions to the problem, can be achieved.

This will, in turn, takes an effect on the procedures which needs to be follow in the research. All matters related to the experimental process and tools will be discussed further in later chapter.

## LITERATURE REVIEW

### 2.1. Translation Technology

Translation technology has been viewed from different perspectives. Alcina

(2008, p. 90), views the research in translation technology as a research which deals with "the design and adaptation of strategies, tools and technological resources that make the translator's job easier as well as facilitating the research and teaching of such activities".

Hutchin and Somers (1992), on the other hand, view the discipline from the involvement of human in using the technology to produce a translation, and from the degree of automation of the translation. Thus, they distinguish this discipline into three (3), fully automated high-quality translation, machine-aided human translation, and human-aided machine translation.

Fully automated translation refers to fully automated translation with a very good quality. This means that no human interference is at all needed in the translation process. Machine-aided human translation defines the process when human are translating with the aid of machine. Human-aided machine translation, in which Youtube's AT and GT are parts of, is the name for the concept when SL is pre-translated and edited by human.

In relation to researches in translation technology, Christensen and Schjoldager (2010), categorize them based on the authors' interests and orientation. There are three categories of the research, they are technology-oriented research, workflow-oriented and industrial research, and translation theoretical research.

Technology-oriented researches focus on the technicalities and evaluation of the technology, or it can be said that in this category, the researches study the technology itself. Some examples of this category is mentioned in Christensen, Flanagan, and Schjoldager (2017) paper which attempts to map the research in translation studies in relation to translation technology from well known journals on translation studies. One research by Lumeras and Way (cited in Christensen, Flanagan & Schjoldager, 2017) points out that statistical based MT will not be replacing human translation any time soon since human has been proven to more superior than the machine when it comes to 'semantic disambiguation, contextual knowledge, and

---

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

genre expertise'. Another study by Carl and Schaeffer (cited in Christensen, Flanagan & Schjoldager, 2017), states that pre-translated literary text is more difficult to decode in terms of syntactic and semantic dissimilarities.

Workflow-oriented research deals with the workflow or process of using the technology. It centers around the technology in action, and other related aspects that influence or being impacted by the use of technology. A more recent study conducted by Garcia (cited in Christensen, Flanagan & Schjoldager, 2017) involving cloud computing reveals that, with it, managers and clients have more control over the work process by 'tapping' into an online resource of casual and inexpensive labor to translate their SL. This research shows the vulnerability of the professionals in the face of technology.

Still in the domain of process-oriented research, Hvelplund (cited in Christensen, Flanagan & Schjoldager, 2017), has conducted an empirical research on the effect of digital resources on the translation process. Her research shows that translators consult the digital resources 20% of the time spent to work on their translation.

In relation to post-editing (PE), Aranberri (cited in Christensen, Flanagan & Schjoldager, 2017) has conducted a research on the quality of MT into a lesser known language. In that research, she concluded that the results of the translation are moderate. What is more interesting, actually, is the PE session conducted by experienced translators on said result. The research shows that the amount of time taken by the translators to reject, revise, or accept it was quite lengthy. This finding is considered to be valuable insight for translator's training and workshops.

Another interesting impact of translation technology is shown in Moorkens and Sasamoto's (cited in Christensen, Flanagan & Schjoldager, 2017) research on translation productivity in relation to the usage and non-usage of technology. They found out that translating from scratch yielded more productivity compared to using translation memory or machine translation, or their variants.

Translation-theoretical research sees translation from the translator's or translation students' point of view, or on the textual aspects of translation. This category also includes the impact of technology on translation training, research, theory, or on the profession of translator.

Research conducted by Kristine Bundgaard, Sharon O'Brien, Maureen Ehrensberger-Dow, Marcel Hasler and Megan Connolly (cited in Christensen, Flanagan & Schjoldager, 2017), all voices the resentment toward the use of technology in their profession.

## 2.2. Evaluation of Machine Translation

In relation to the evaluation of MT, Hutchins and Somers (1992), has pointed out that by then, the evaluation is based on performance, and not on the potential. They further assert that the research was conducted by non-practitioner of MT. In addition, those which were made by MT practitioner are 'minimal and misleading.' Unfortunately, there is no mention of why they are like that.

They further add that the evaluation should be done on different stages of the machine development. The first evaluation, according to them, should be performed during the development of the prototype. This stage is called the prototype evaluations.

The second stage of the evaluation is named development evaluation. In this stage, evaluation is done to the system performance in the intended environment. The third stage is related to the users of the system. In this stage of evaluation, the amount of human contribution in the post editing becomes one of the important factors. In addition, this stage also includes the evaluation of the required working condition and requirement, the compatibility of the system with existing equipment, and the required technical facilities. This stage is also called as translator's evaluation.

The final stage is classified as the recipient's evaluation. In this stage, the quality of the result is being put under the microscope.

Closely related to the field of linguistics, they also state that in every stage,

---

\*Author(s) Correspondence:

E-mail: [ataufik@bundamulia.ac.id](mailto:ataufik@bundamulia.ac.id)

there are linguistics evaluations. The linguistics evaluation is further divided into the evaluation of quality and the identification of errors.

### 2.2.1. Evaluations of Quality

The evaluation of quality includes the the evaluation of accuracy, clarity, and style. In the evaluation of accuracy, several proposals have been put forward, from the testing of the translation's informativeness, which is very subjective, to the conformity of performance, and back translation.

In terms of clarity, the use of readability scales, cloze technique, and comprehension tests have been conducted to test it. In relation to style, the subjectivity is too high, so, in their book, there has been no mentions about the proposed techniques to evaluate it.

### 2.2.2. Identification of Errors

The method which has been used in this case is the counting of error. What is meant by error in this case is the amount of addition, deletion, and substitution needed for the post editor to create an acceptable translation. To add the validity to the evaluation based on errors, however, a simple error counting will not be sufficient; classification of errors and the correction difficulty should also be included. Finally, there is also a benchmark test to identify the errors. This is conducted by comparing the result of the MT with a corpus of text. However, the obvious problem would be to find out the text which encompasses the linguistics phenomena which become the center of the problem, the involvement of the user in post-editing process, and many other factors (Hutchins & Somers, 1992).

### 2.3. Google Translate

There have been notable researches focusing on the quality of GT. One notable experiment which can be exemplified in the testing of the accuracy is the testing of understanding. In his article, 'The Shallowness of Google Translate' Hofstadter (2018) discovered that a linguistics element, namely

gender specific expression, such as one found in French, cannot be identified accurately by GT. This experiment is conducted from both English to French, and vice versa, and it is tested on the 'new and improved' neural based system of GT which is said to be enhanced by 'Deep Learning'.

Hofstadter also criticizes GT's choice of word. Since GT is based on statistics, it chooses the most commonly occurring equivalence to assign to a word. It does not, or rather, is not able yet, to go beyond their database of equivalence. Hence, the word choice often cannot agree to the situation behind the word used in the SL.

Simply put, in this very recent research or experiment (2018), Hofstadter concluded that what GT is doing, at the moment, is only relating word with other word; there isn't any indication yet, even with its neural system, which is supposed to mimic brain function, to integrate the element of ideas, experience, images, or even memories into its system, which not only are the elements that our mind or brain can do, but also very crucial in translation.

On a side note, however, the language pairs used by him was not Indonesian. So, it is interesting to see whether the same will happen to the English-Indonesian pair. In addition, the researcher feels the need to emphasize once again that finding errors in GT is not the researchers' intention, but rather what can be done to minimize the problems caused by the errors, or at least, minimize the post-editing process.

### 2.4. Equivalence

In section 2.2, it has been mentioned that the evaluation of machine translation can be done by looking at the naturalness, accuracy, and clarity of the text. It has also been mentioned that the three criteria can be assessed using translators' interpretation and assessment of the text, which is very subjective. Another method to assess the quality of the text is through the use of error identification. This was mentioned in section 2.3. This, on the other hand, is polar opposite to the previously mentioned method. In it, errors are quantified

---

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

and categorized into addition, substitution, and deletion. There was also mentions that corpora can further help with the validity of the assessment.

One notable missing aspect from both methods is the analysis of the text itself. Therefore, the researcher has decided that a framework for analysis needs to be proposed. For that purpose, the researchers have turned their attention toward the concept of equivalence.

For this, the researcher turns to Mona Baker and her excellent book, *In Other Words: A Coursebook on Translation* (1992). In it, she discusses the types of equivalence. The first one is word level equivalence. What is meant by word in her book is the smallest element of language which has meaning, and that is morpheme. In relation to meaning, she differentiated four types of meaning: Propositional, evoked, presupposed, and expressive meaning.

Propositional refers to the actual meaning of a word, while expressive meaning relates to the speaker's feeling. Presupposed meaning is about meaning in a collocation. The meaning of words in collocation are restricted by the word it collocates with. The restriction can be further categorized into two; selection and collocation. Selection restriction often happens with words with propositional meaning. Collocation restriction, on the other hand, refers to the commonly co-occurring words. In other word, the word is restricted by the common collocation of a specific situation or geographical area.

Evoked meaning is related to dialect and register. The cause of dialect is further categorized into three, geographical, social class, and temporal. Geographical dialect is related to the dialect differences caused by different geographical location. Dialect can also be different due to the different social class the speakers have. Finally, Temporal dialect refers to differences in dialect because of age difference.

Other factor influencing evoked meaning is the register which consists of field, tenor, and mode. Field decides the people's linguistics choices based on the condition in

which they have to use the language for. Tenor takes into considerations the people taking part in the discourse. Mode is related to the medium used to communicate; it can be spoken or written.

Based on the previous definitions, she has made the common problem of equivalence at word level.

The first one is culture specific concept (CSC). As its name says, it is related to the items which are culture-specific. The second problem is about the words which are not lexicalized in the target language. This means that the problematic words are specific to the source language although they are not culture-specific. The third problem concerns the complexity of the source language. Sometimes, even when a word has its propositional meaning, there is another set of meanings which entail it.

The next problem is related to the different distinction made by the source and target language. What one language consider different in meaning, is considered similar in another language. A lack of superordinate and hyponyms also creates a problem in producing equivalent. One language might have only one verb in expressing one type of action. Another language might have more than one verbs to express the same action. Usually this is related to expressive meaning.

Differences in physical or interpersonal perspective sometimes create a problem. English has words for different physical settings like come and go. Some languages however, does not have that and instead have made the difference in meaning based on the interpersonal perspective.

Differences in form, frequency and purpose of using specific forms also create trouble in finding equivalence. Certain suffixes in English, like the suffixes indicating plurality do not always have equivalences in other language. Similarly, the frequency of use and the purpose of using a specific form in one language is not always transferred flawlessly into other language. Loan words also become a problem in finding equivalence. Sometimes it can be very tricky indeed to give equivalence for this since there is a possibility of using

---

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



false friend; word/s which have similar form but different meaning.

Other than proposing the common problems of equivalence at word level, Baker also suggested different strategies in coping or dealing with them. The first strategy is to use a more general word. This is usually employed when we are dealing with words with propositional meaning which are not lexicalized equivalently in the target language.

The second strategy is to translate using a more neutral or less expressive words. As can be guessed from the name, it can be used to deal with expressive meanings. It certainly eliminates the expressiveness of the target language, however, sometimes it can be preserved using modifiers.

Cultural substitution can be used when translating cultural specific items. Audience can be made to know the significance of the culture in the source language if it is substituted by well-know cultural items of the target language.

One way to deal with equivalence is to loan the words which become the focus, or loan them with explanation. If the translator thinks that the audience will know the meaning of the loan words, than just borrowing it should be enough. However, in the case that loan word is not enough, explanations can be given to clarify it.

Another strategy which could help with equivalence is to paraphrase it. There are two ways to do this. One, is by using related words. This is usually done if the concept is lexicalized in the target language, only in different form. The second method of paraphrase is to use unrelated words. This is obviously done when the concept is not lexicalized in the target language. There are disadvantages of using these forms of paraphrase. The first one is that, by paraphrasing the concept, it will lose its expressive meaning. The second is the substitution of item into items. This could be very problematic if space is important, like in audiovisual translation.

Omitting the unnecessary words can also be a solution. It is possible to preserve the lexical meaning. Finally, there is the use of

illustration. This is used if illustration is allowed in the text. This is very advantageous if space is important.

As stated earlier in this chapter, a framework of analysis is needed in analysing the product of GT and YouTube translation. The researcher will use this proposal by Baker as his framework of analysis. The common problem of equivalence will be the basis of choosing the problem area, while the strategies and the categorizations of lexical meanings will be used to analyze the products of translation by GT and YouTube Translation of the problem area.

## RESEARCH METHODOLOGY

### 3.1. Ontology and Epistemology

In their very loose, and generic definition, Matthews and Ross (2010) defines that ontology is the way we see the social world and the 'nature and reality of social phenomena which makes up the social world'. Epistemology, on the other hand, is defined (Matthews & Ross, 2010) as the theory of knowledge, or how we know things.

Furthermore, Matthews and Ross (2010) distinguish ontology into three categories; objectivism, constructivism, and realism. They also state that the three different ontological categories have their own theory of knowledge, namely positivism, interpretism, and realism.

Objectivism sees social phenomenon as a self-existing being; claiming that no social actors contribute to its existence. Constructivism, on the other hand, sees social actors as active contributor to the existence of a social phenomenon. Realism believes that social phenomenon can exist without social actors, but als believes that social actors can be involved in the creation of social phenomenon.

The positivist epistemology adopts the ontological point of view of objectivism. Positivists believe that social phenomenon can be objectively researched, and the data acquired from the social world can be calculated, measured, and free of subjectivism. Interpretism assumes constructivistic ontological position. Interpretism allows social

---

\*Author(s) Correspondence:

E-mail: [ataufik@bundamulia.ac.id](mailto:ataufik@bundamulia.ac.id)

phenomenon to be interpreted from the point of view of its actors and their subjective interpretation. It is commonly associated with qualitative approaches. Realist researcher adopts both the ontological and epistemological position of realism. In short, it is a combination of qualitative and quantitative approaches.

Having seen the generic explanation of ontology and epistemology, it is now time to decide the approach which must be taken to take this research on. This research aims to identify the problems which exist in the translation results of the two systems, and to propose possible solutions to the problem. The accuracy or the quality of Youtube's AT and GT are not measurable by numbers. The evaluation later will be based on the 'agent' interpretation of the translation product, and a triangulation of the interpretation with a set of corpora. Having seen the nature and reality of the social phenomenon, it is safe to assume that this research assumes the ontological position of realism epistemology and ontology. This position is often related to a mix of qualitative and quantitative approaches (Saldanha and O'Brien, 2014).

### 3.2. Research Methodology

#### 3.2.1. Model of Research in Translation Studies

In a research there are a lot of terminologies which often overlap. To avoid that, this paper will adopt the definitions proposed by Saldanha and O'Brien (2014). First of all, the research will clarify what is meant by model. Saldanha and O'Brien has distinguished research models into four (4), they are:

- 1) Product-oriented model
- 2) Process-oriented model
- 3) Context-oriented model, and
- 4) Participant-oriented model

Product-oriented model is similar, in a sense to Marco's textual descriptive model (2009), or Chesterman's comparative model (2000). Product oriented model of translation studies, as the name suggests, focus on the product of translation. This research on

product can use descriptive/exploratory and evaluative approaches.

Descriptive or explanatory approaches in translation studies is known to use Critical Discourse Analysis (CDA) and Corpus Linguistics (CL).

CDA sees discourse as a social practice (Fairclough, 2003). In his analysis, Fairclough proposes a model of CDA which has three dimensions (text, discursive, and social practice). Moreover, in his analytical approach, CDA sees language as bearer of change, tool for power, and behavior modification. Because of this highly interpretism standing, all CDA practice did not adopt objectivism and positivism ontology and epistemology respectively.

CL centers around corpora. Corpora is a set of authentic text which is categorized based on specific criteria (Bowker & Pearson, 2002). In some ways, CL is similar to CDA in terms that it needs the whole text in the investigation, the link between text and context, as well as the use of naturally occurring language. However, CL, with its heavy emphasis on frequency, might not be compatible with interpretism ideology.

There are some weaknesses of using CDA such as argument circularity, privileged knowledge, and partiality. However, there has been suggestion to overcoming the weaknesses by making the data collection and analysis more explicit, and setting criteria for selecting the text, and triangulating the result with data obtained using other method. Speaking about triangulation of data, CL has been used frequently to triangulate the claims of CDA. In turn, in terms of detailed description, CDA will provide assistance to CL. So, the combination of the two is the ideal usage. The combination of both qualitative and quantitative approaches shows approval toward realism ontology and epistemology.

Process oriented model, or process research, focuses on the translators; especially on their expertise, competence, and the cognitive process behind it. This process-oriented research is similar to Chesterman's process model and Marco's cognitive model. There has been numerous arguments that a

---

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

process research should be combined with other model to provide richer results. Some which have been conducted are the combination between process research and product research and corpus based research. Since triangulation is needed in process research to increase its validity, qualitative and quantitative methodology is needed to produce a better understanding of the process. Hence, this type of research belongs to the realism ontology and epistemology.

The participant-oriented research incorporates the concept or theory from other discipline in order to understand the translation. Looking at its dependence toward other discipline's theory and concept, and from the methods used to collect the data (questionnaire, focus groups, and interview), this type of research agrees with the interpretism epistemology. However, it can also be made empirical, if the research questions allow it to.

Context-oriented research focuses on the external factors influencing the translator, the situation in which the translation takes place, and the way translation affects culture. Context oriented model is similar to Marco's culturalist and sociological model, or Chesterman's causal model. Case studies are frequently used to explain a phenomenon in a context-oriented research. Data in case studies can be collected from written sources, interviews, observation, physical artefacts, and quantitative data.

Looking at the models explain previously, in which model does translation technology fall into? To answer that question, let us look back at the categorization or themes set up by Christensen and Schjoldager (2010): Technology-oriented research, workflow-oriented and industrial research, and translation theoretical research.

The examples included in chapter two that belong to the first category focuses on the evaluation of the technology. The evaluation of the technology is conducted on the basis of the result of the technology itself. From this, it can be said that the technology-oriented research can be included in the product oriented research. Therefore, the methods used

can also use the methods used in said models, namely CDA or CL, or the combination of the two.

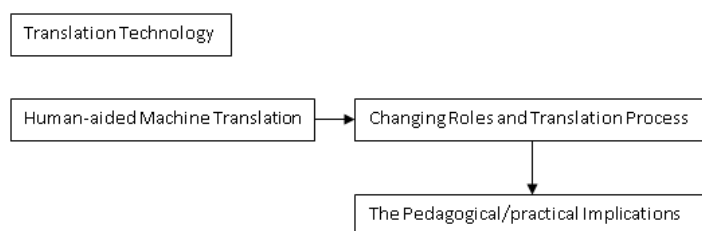
The workflow-oriented and industrial research is similar to the Saldanha and O'Brien's process and context-oriented model or Marco's sociological and culturalist model. The examples shown in the previous chapter are focused on the process of using the technology and how technology affects the socio-cultural status of translators.

The examples shown in the theme of translation-theoretical research show the translator's attitude toward technology. This thematic categorization of research can then be concluded to belong to the epistemological standing of interpretism. So, based on these themes, it is safe to say that this current research falls into the thematic distinction of technology-oriented research and product-oriented model, since the focus of the research is to assess the result of translation produced by the technology by conducting an in depth textual analysis on the results and the triangulate them using a set of corpora.

### 3.3. Research Roadmap

Having finished the last roadmap on automated speech recognition, the author now turns his attention toward translation technology and computational linguistics. This research on the quality of HMT will be the starting point for further research which includes the changing roles of translator as a result of the introduction of technology, and its pedagogical and practical implications. The following illustrates the research roadmap.

**Figure 1. Roadmap on Translation Technology**



\*Author(s) Correspondence:

E-mail: [ataufik@bundamulia.ac.id](mailto:ataufik@bundamulia.ac.id)



## DATA ANALYSIS

### 4.1. Data Analysis

As stated in Chapter 2 of this research, there are several methods which can be used to choose the problem area. This research is using the common problem of equivalence at word level defined by Baker (1992).

The first problem is related to culture specific concept (CSC). According to Baker,

the concepts included in culture specific concepts are usually related to food, religious beliefs, and social customs which are totally unknown in the target language. As per this definition, the researchers have collected thirty (30) words which are related to said categories, they are:

**Table 1. List of Indonesian Culture Specific Concepts**

No.	CSC – Social Customs	CSC – Religious Beliefs	CSC – Types of Food
1	Kebo-keboan	Alkitab	Otak-otak
2	Tiwah	Ayat	Empek-empek
3	Tabuik	Hadis	Kerak Telor
4	Adu Betis	Nirwana	Gudeg
5	Dugderan	Wihara	Papeda
6	Bakar Tongkang	Karma	Rendang
7	Batombe	Adventis	Gulai Belacan
8	Brobosan	Roh Kudus	Surabi
9	Tatung	Katekismus	Bubur Pedas Sambas
10	Gigi Runcing	Purgatorium	Kepiting Soka

Source: Taken from various sources

And here is the translation of the concepts in context using GT. The translator is using context in presenting since it was also

done in the examples provided by Baker in her book.

**Table 1. GT Results of CSC Concepts: Social Customs (in context)**

No.	Indonesian	English (GT)
1.	<b>Kebo-Keboan</b> adalah salah satu upacara yang diselenggarakan masyarakat Banyuwangi.	<b>Kebo-Keboan</b> is one of the ceremony held by Banyuwangi community.
2.	<b>TIWAH</b> merupakan nama sebuah upacara penghantar kerangka jenazah ke tempat peristirahatan yang terakhir	<b>TIWAH</b> is the name of a ceremonial beautification ceremony to the last resting place
3.	Festival <b>Tabuik</b> merupakan salah satu tradisi tahunan di dalam masyarakat Pariaman	<b>Tabuik</b> Festival is one of the annual traditions in Pariaman society
4.	Tradisi <b>adu betis</b> ini biasanya diadakan sebagai bentuk rasa syukur atas keberhasilan panen.	The tradition of <b>calf fighting</b> is usually held as a form of gratitude for the success of the harvest.

5.	<b>Dugderan</b> merupakan festival untuk menandai dimulainya ibadah puasa di bulan Ramadan yang diadakan di Kota Semarang	<b>Dugderan</b> is a festival to mark the commencement of fasting in Ramadan held in Semarang City
6.	Ritual <b>Bakar Tongkang</b> dikenal juga sebagai Upacara Bakar Tongkang	<b>Tongkang Bakar</b> Ritual is also known as <b>Barge Burning Ceremony</b>
7.	<b>Batombe</b> adalah tradisi berbalas pantun yang berasal dari Nagari Abai, Kecamatan Sangir Batang Hari, Kabupaten Solok Selatan, Provinsi Sumatera Barat.	<b>Batombe</b> is a trademark rhyme that comes from Nagari Abai, Sangir Batang Hari Subdistrict, South Solok Regency, West Sumatra Province.
8.	Termasuk di dalamnya pula adalah tradisi kematian bernama <b>Brobosan</b>	Included also is the tradition of death named <b>Brobosan</b>
9.	<b>Tatung</b> dalam bahasa Hakka adalah orang yang dirasuki roh dewa atau leluhur	<b>Tatung</b> in Hakka is a man possessed by the spirit of a god or an ancestor
10.	Selain itu, tradisi <b>gigi runcing</b> ini juga dipercaya akan bisa memberikan kedamaian jiwa pada seorang wanita	In addition, the tradition of <b>pointed teeth</b> is also believed to be able to provide peace of mind to a woman

Source: Context are taken from various sources

As per Baker's explanation, when dealing with cultural specific items, translators can make use of cultural substitution. However, in the case of GT, it was found that loan word and related-word paraphrase is chosen. From the 10 (ten) items related to social custom, 7 (seven) are translated using loan words, and 3 (three) others are translated using paraphrase of related concept.

In terms of acceptability, the researchers triangulate them with the entries in COCA (Corpus of Contemporary American English). However, after triangulating the results of GT into COCA, no matches were found. The reserachers therefore were forced to triangulate them using search engine.

The result of the triangulation of the words without any context attached was disappointing. This is because there were very little matches found. However, after trying different method of triangulation, the researchers were quite surprised with the new results. That new way is to use the context of the words. For example, rather than just put the word 'dugderan', the researchers include the word 'festival' to collocate with it. More matches are found using this method, and they were related to the words and contexts being

discussed. This strategy, according to Baker, is named loan word plus explanation.

One entry suggested in this paper, however, has been translated inappropriately. That involves the word 'gigi runcing'. GT translated it into 'pointed teeth'. After the triangulation with the search engine, it was found that the acceptable translation is actually 'chiseled teeth'.

With those findings, the researchers can conclude the followings:

- a) Baker's strategy of using cultural substitution might not be applicable in Indonesian context since there are very unique social customs in Indonesia which might not have cultural equivalent anywhere else in the world;
- b) Instead, the appropriate strategies in translation Indonesia's social customs are to use loan word plus explanation and to paraphrase the customs or traditions using related words.
- c) GT have produced an acceptable rendition of the Indonesian social custom into English. Despite the need of context, 90% of the translation of Indonesian social custom by GT is

\*Author(s) Correspondence:

E-mail: [ataufik@bundamulia.ac.id](mailto:ataufik@bundamulia.ac.id)

acceptable. This is proven via the triangulation with the search engine. Unfortunately, triangulation with corpus did not reveal any satisfactory results. There are a lot of reasons for

this. One of which is the lack of entries in the corpus itself.

Similar results are found when the researcher used YT to translate the sentences. The examples are seen below.

**Figure 2. Example 1 of YT Results**



**Figure 3. Example 2 of YT Results**



Figure 4. Example 3 of YT results



Table 3. GT Results of CSC Concepts: Religious Beliefs (in context)

No.	Indonesian	English (GT)
1.	Banyak orang yang meragukan kebenaran <b>Alkitab</b> karena berbagai-bagai alasan.	Many people doubt the truth of the <b>Bible</b> for various reasons
2.	<b>Ayat Al-Quran</b> mana yang sesuai denganmu tahun 2018?	Which <b>verses</b> of the <b>Quran</b> correspond to you in 2018?
3.	Secara struktur <b>hadis</b> terdiri atas dua komponen utama yakni <b>sanad/isnad</b> (rantai penutur) dan <b>matan</b> (redaksi).	The structure of <b>hadith</b> consists of two main components namely <b>sanad / isnad</b> (chain of speakers) and <b>matan</b> (redaction)
4.	Saya mohon maaf kalau ada tulisan-tulisan saya sebelumnya yang kesannya surga dan <b>Nirwana</b> adalah sama.	I apologize if any of my previous writings that the impression of heaven and <b>Nirvana</b> are the same
5.	<b>Wihara</b> adalah rumah ibadah agama Buddha, bisa juga dinamakan kuil.	The <b>temple</b> is a Buddhist temple, it can also be called a temple.
6.	<b>Karma</b> kerap kali diartikan sebagai kejadian yang buruk. Namun hal itu juga berlaku bagi balasan baik atas perbuatan yang dilakukan.	<b>Karma</b> is often interpreted as a bad event. But it also applies to a good reply to the deed done.
7.	Ada beberapa orang tokoh yang dapat disebut sebagai pelopor/pendiri aliran <b>Adventis</b> ini.	There are some prominent figures who can be called the pioneers / founders of this <b>Adventist</b> school.
8.	<b>Roh Kudus</b> adalah pribadi Tuhan dalam konsep <b>Tritunggal</b> .	The <b>Holy Spirit</b> is the person of God in the concept of the <b>Trinity</b> .
9.	<b>Katekismus</b> adalah suatu ringkasan atau uraian dari doktrin yang umum	The <b>Catechism</b> is a summary or description of the doctrine commonly

\*Author(s) Correspondence:

E-mail: [ataufik@bundamulia.ac.id](mailto:ataufik@bundamulia.ac.id)

	digunakan dalam pengajaran agama Kristen (katekisasi), baik untuk anak-anak maupun orang dewasa.	used in the teaching of Christianity (catechism), both for children and adults.
10.	Istilah bahasa Latin <b>purgatorium</b> atau dalam bahasa Inggris purgatory diterjemahkan dengan api penyucian.	The Latin term <b>purgatory</b> or in English purgatory is translated by purgatory.

Source: Context are taken from various sources

From the results, it can be seen that the strategy used to CSC which are related to religious concept is in accordance to what Baker has stated, that is cultural substitution. This is evident in all words related to the top religions in world, namely Christian, Islam, Buddha, and Hindu. There are however, several points of discussion from it.

The first concerns the words taken from Islam, which are 'verse' and 'hadith'. After triangulating the results with COCA, it was found that those two words were last seen in the form of written in 2016. However, a further cross-referencing reveals that if it is spoken, such words are still used. This is proven by conducting a search on those words in YouTube. Both words are still in use in 2018. The use of the word 'verse' which is related to Islam can still be seen from a video in YouTube in the channel *ThereIsNoClash*.

Similarly, the word 'hadith', which was last seen on the written form in 2016, according to COCA, in a magazine article published by Huffington Post, can still be

found in a YouTube channel named *TheProphetsPath*.

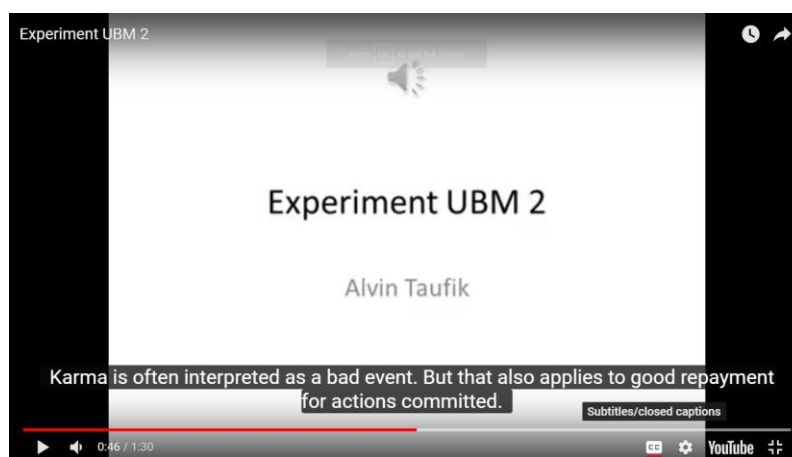
Similar finding is also visible in the word 'nirvana'. In COCA, that word was last seen in 2014. However, a YouTube channel named *Positive Magazine Meditation* still use that word in its video. The true meaning of that word is also referred to in the video.

To sum up, the followings contain what the researchers think about the results of GT on CSC related to religious beliefs.

- In translating religious-related CSC, GT has done a decent job. This is proven from the triangulation to COCA and YouTube
- Baker's cultural substitution is the proper strategy to be used in this case, and at surface level, GT has already adopt that. A more complex research needs to be done to prove whether that statement is true or not.

Similar results are found in the translation using YouTube.

Figure 5. Example 4 of YT Result



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



The third and final experiment is conducted on the final type of CSC, namely

food types. The following is the table for the result of GT.

**Table 4. GT Results of CSC Concepts: Types of Food (in context)**

No.	Indonesian	English (GT)
1.	Jika anda tertarik untuk berbisnis frozen food, <b>otak-otak</b> dapat menjadi salah satu produk andalan.	If you are interested in frozen food business, the <b>brain</b> can be one of the mainstay products
2.	Nama pempek sendiri diambil dari kata <b>empek-empek</b> , atau pek-pek yang menjadi sebutan laki-laki tua pada pria Tionghoa.	Pempek name itself is taken from the word <b>empek-empek</b> , or pek-pek that became the old male title in Chinese men
3.	<b>Kerak Telor</b> menjadi salah satu kuliner khas warga Jakarta.	<b>Egg crust</b> is one of the typical culinary of Jakarta residents
4.	<b>Gudeg</b> telah dikenal oleh masyarakat Indonesia khususnya sebagai makanan khas dari Kota Yogyakarta.	<b>Gudeg</b> has been known by the people of Indonesia, especially as a typical food from Yogyakarta.
5.	<b>Papeda</b> adalah makanan berupa bubur sago khas Maluku dan Papua yang biasanya disajikan dengan ikan tongkol atau mubara yang dibumbui dengan kunyit.	<b>Papeda</b> is a food in the form of Maluku and Papua sago porridge which is usually served with tuna or mubara fish seasoned with turmeric.
6.	<b>Rendang</b> atau randang adalah masakan daging bercita rasa pedas yang menggunakan campuran dari berbagai bumbu dan rempah-rempah.	<b>Rendang</b> or Randang is a spicy flavored meat dish that uses a mixture of various herbs and spices.
7.	Salah satu masakan yang sedap dan mantap untuk disantap adalah <b>gulai belacan</b> khas Riau	One of the delicious and steady dishes to eat is typical <b>Riau curry curry</b> .
8.	Satu di antara jajanan tradisional di Bandung yang patut traveler coba adalah <b>surabi</b> .	One of the traditional snacks in Bandung that is worthy of a traveler try is <b>surabi</b> .
9.	<b>Bubur pedas</b> merupakan masakan khas suku Melayu Sambas di Kalimantan Barat.	<b>Spicy porridge</b> is a typical Balinese Sambas tribe in West Kalimantan.
10.	Budidaya <b>kepiting soka</b> adalah salah satu peluang usaha yang dapat mendatangkan untung besar.	<b>Soka crab</b> cultivation is one business opportunity that can bring big profits.

From the result of GT, we can see that the food-related-CSC is translated using different strategies. The first one is cultural substitution. This strategy is used in the word 'bubur pedas', which was translated into 'spicy porridge'. The writer notes that even though it may refer to different concept,

this translation is appropriate since it explains clearly how the porridge tastes.

The second form of the translation is the use of loan word. This can be found in the words 'Gudeg', 'papeda', 'empek-empek', 'rendang', and 'surabi'. There are, however, different interpretation which can be made

\*Author(s) Correspondence:

E-mail: [ataufik@bundamulia.ac.id](mailto:ataufik@bundamulia.ac.id)

after they were triangulated with different platform. The ones which are ‘appropriate’ were the ‘translation’ of the words ‘rendang’ and ‘gudeg’ since they were both found in COCA and they refer to similar concepts. As for the words ‘empek-empek’, ‘surabi’, and ‘papeda’, they cannot be found in COCA. However, for the words ‘surabi’ and ‘papeda’, they can be found in Google and YouTube respectively. Take notes, that when the researchers are looking for a match in both Google and YouTube, the researchers focus on a match taken from foreign concepts, not the Indonesian ones. In addition, the word ‘empek-empek’ can only be found in Indonesian concept.

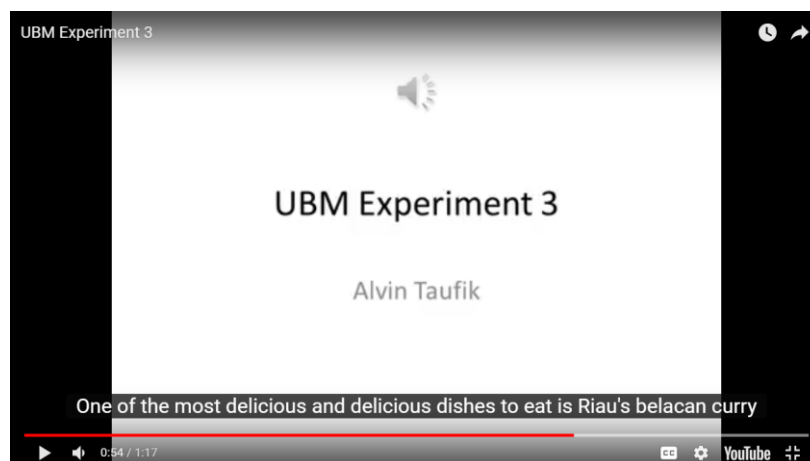
The last form of translation for food-related-CSC is the use of paraphrases using related words. This can be found in the words/phrase ‘egg crust’, ‘Riau curry’, ‘brain’ and ‘Soka crab’. As with the previous form, different interpretation can also be made on these. First of all, the researchers would like to point out the obvious, which is the translation of ‘otak-otak’ into ‘brain’. The researchers did not feel the need to triangulate this into the three platforms used earlier since we believe that it will refer to a larger semantic group than the SL. This is also true with the phrase ‘Riau curry’. The word ‘curry’ generates a wider concept than the one referred to by the food. The second is on the translation of ‘kepiting soka’ into ‘soka crab’. Further

research has shown that ‘soka’ is supposed to be translated into ‘soft shell’. So, it can be said that the translation does not provide clear explanation of the object being discussed. Finally, with the phrase ‘egg crust’, we were again faced with the case of different concept. However, unlike the phrase ‘spicy porridge’, the ‘egg crust’ of the SL does not have ANY similarities with the TL concept. Thus, it can be said that the cultural substitution for this phrase was not appropriate. Having discussed the result of the translation, the following conclusions can be made:

- a) Unless having similarities with the SL, cultural substitution should not be used in translating food-related-CSC.
- b) For those lesser known food-related-CSC, paraphrases with related words might be better used.
- c) As with social customs, the better strategy to be used in the translation of food-related-CSC is the paraphrases using related words and loan words.

Similar results can also be found in the YT version of the translation, with some minor exceptions. One of the few exceptions is the additional loan word in the translation of the phrase ‘gulai belacan’ which was translated into ‘Riau’s belacan curry’. However, the researchers do not think of this as something significant as it is still ‘going to be foreign in the TL concept. Here is the example of it.

Figure 6. Example 4 of YT Result



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

## CONCLUSION

### 5.1. Conclusions

The aims of this research is to identify the problems with GT and YT and to propose a solution for it. From this research it was found that most of the translation for the Cultural Specific Concepts conducted by GT and YT are in the form of loan words. Although this is not in accordance to the theory, after triangulation, it was found that this is the most appropriate form of translation, especially for highly popular words.

Cultural substitution, which is suggested by Baker, is also used to translate the CSC. This form is appropriate for several CSC since the TL has similar concept, or at least significant similarities with TL concept.

If there is a solution which can be given to the translation provided by GT and YT, it is that the translations which use Loan Word, especially on those which is unknown in the TL context, paraphrases using related words or loan words with explanation should be applied. This suggestion or solution, is directed of course to the user of GT and YT since there are no possibilities that the user can modify the results of the translation.

## REFERENCES

- Alcina, A. (2008). Translation technologies: Scope, tools and resources. *Target*, 20(1), 79-102.
- Baker, M. (1992). *In other words: A coursebook on translation*. London & New York: Routledge.
- Bowker, L., & Pearson, J. (2002) *Working with specialized language: A practical guide to using corpora*. London & New York: Routledge.
- Chesterman, A. (2000). A causal model for translation studies. In M. Olohan (Ed.), *Intercultural faultlines: Research models in translation studies I – textual and cognitive aspects* (pp. 15-28). Manchester: St. Jerome.
- Christensen, T. P., & Schjoldager, A. (2010). Translation memory (TM) research: What do we know and how do we know it?. *Hermes – Journal of Language and Communication Studies*, 44, 89-102.
- Christensen, T.P., Flanagan, M., & Schjoldager, A. (2017). Mapping translation technology research in translation studies: An introduction to the thematic section. *Hermes – Journal of Language and Communication Studies*, 56, 7-20.
- Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research*. New York: Routledge.
- Hofstadter, D. (2018, January 30). *The shallowness of Google Translate*. Retrieved from <https://www.theatlantic.com/technology/archive/2018/01/the-shallowness-of-google-translate/551570/>.
- Hutchins, W. J., & Somers, H. L. (1992). *An introduction to machine translation*. London: Harcourt Brace Jovanovich.
- Marco, J. (2009). Training translation researchers: An approach based on models and best practice. *The Interpreter and Translator Trainer* 3(1), 13-37.
- Matthews, B., & Ross, L. (2010) *Research methods: A practical guide for the social sciences*. Edinburgh: Pearson Education Ltd.
- Saldanha, G., & O'Brien, S. (2014). *Research methodologies in translation studies*. New York: Routledge.

\*Author(s) Correspondence:

E-mail: [ataufik@bundamulia.ac.id](mailto:ataufik@bundamulia.ac.id)