Language Literacy: Journal of Linguistics, Literature, and Language Teaching
Volume 6, Number 1, pp: 16-27, June 2022
e-ISSN: 2580-9962 | p-ISSN: 2580-8672
DOI: 10.30743/ll.v6i1.4678

PROPOSITION-BASED EVALUATION OF MACHINE-TRANSLATED ACADEMIC TEXT

Aulia Addinillah Arum, B. R. Suryo Baskoro
Universitas Gadjah Mada, Yogyakarta, Indonesia
E-mail: aulia.addinillah@mail.ugm.ac.id

Received: 2021-12-16 Accepted: 2022-01-24 Published: 2022-06-28

Abstract

This study aims to describe the rendition of propositional meaning in machine-translated academic text. A proposition is that part of the meaning of a clause or sentence that is constant, despite changes in such things as the voice or illocutionary force of the clause. A proposition may be related to other units of its kind through interpropositional relations, such as temporal relations and logical relations. To assess whether the meaning of an utterance is conveyed adequately in the target text, we conducted the proposition-based evaluation by looking at the grammatical structure, semantic roles, and the category of proposition reflected in the source text and the target text. The analysis is done by adopting the qualitative approach based on Larson's theory of Meaning-Based Translation. The findings of this study suggest that identical grammatical structure can have a positive correlation to the semantic structure and the transfer of meaning in machine translation. This study also reveals that grammatical-structure similarity does not always indicate meaning accuracy in translation.

Keywords: *Translation Evaluation; Machine Translation; Translation Quality; Meaning; Semantic Role*.

1. Introduction

There will be no single objection to the claim that, in most text types, a good translation must prioritize the delivery of meaning rather than the transference of form. This claim is based on the fact that meaning, in whatever the lexical unit is (words, phrases, sentences, etc.), is always "packaged" in a different form when rendered in a different language. Therefore, the task of a translator is to preserve the meaning from the source language (SL) into the target language (TL) (Larson, 1998; Wu & Xu, 2011).

Nevertheless, the skewing between meaning as the deep structure and form as the surface structure of an utterance or text brings a significant challenge to translators in delivering the intended meaning of the (SL) with a natural form. For instance, in a simple sentence, such as "*The building is poorly designed"*, the word *building* is categorized as a noun since it is positioned as a subject at the surface structure. However, semantically, *building* can also refer to an action representing an Event in the semantic category. Another consideration a translator must take is the distinction between the primary meaning and secondary meaning of a lexical item and the primary and secondary functions of grammatical markers or is also referred to as the deep structure and the surface structure of meaning (Chomsky, 2019; Larson, 1998; Wu & Xu, 2011).

Given that meaning becomes the utmost priority in translation, therefore, there should also be a radical shift in evaluating the quality of a translation product, i.e., from form-based evaluation into meaning-based evaluation. This new evaluation approach should not be limited to evaluating human translation (HT), but should also be applied in the evaluation of machine translation (MT). Within the computational domain, the machine translation output is often evaluated with automatic evaluation methods that involve quantitative measurement. BLEU (BiLingual Evaluation Understudy) is one of the available automatic evaluation tools which rely upon lexical similarity principle between the text under examination (candidate text) and the "gold standard" translation (human translation/reference text) (Papineni et al., 2002). BLEU evaluation compares the n-gram matches between each candidate translation and the reference translations (Koehn, 2010; Papineni et al., 2002).

Despite the advantages offered by automatic evaluation tools, which are perceived to be less expensive and able to deliver immediate results (Papineni et al., 2002), the use of these tools also brings some shortcomings, particularly when its performance is viewed from a linguistic perspective. The major pitfall in using automatic evaluation is a very limited quality assumption, given that this approach mainly relies upon lexical similarities (Giménez & Màrquez, 2010). Furthermore, automatic evaluation has not had the ability to capture meaning above words or phrase level. Therefore, to conduct a comprehensive evaluation, we must start by determining the right unit of analysis. In this study, we would argue that proposition is the most comprehensive basis for the manual evaluation since it can assess the overarching meaning of a text rather than just assessing the words-to-words correspondence.

However, our aim is not to display the superiority of manual evaluation over automatic one. We truly believe that the automatic evaluation is utterly beneficial to the development of automatic translation systems, especially in terms of providing a quantitative measurement of grammatical and word-level equivalence. Instead, our goal is to reveal the potential use of proposition in the manual of evaluation of machine translation output as a complementary approach in conducting a comprehensive MT evaluation. This exploratory study presents our findings based on the following research questions: 1) How is the quality of MT assessed through proposition-based manual evaluation? 2) What are the major causes of poor rendition of proposition from SL to TL?

2. Literature Review

2.1 Proposition

In principle, a proposition is an idea unit (DeFrancesco & Perkins, 2012). Within the domain of language philosophy, proposition can be defined as an abstract entity that has three roles: 1) to be the meaning of a sentence (at a context of utterance); 2) the object of propositional attitudes; and 3) the vehicles of truth and falsity (Kemp, 2018). In addition, Fillmore (2020) also defined *proposition* as a tenseless set of relationships involving verbs and nouns (and embedded sentences, if there are any), separated from what might be called modality constituent. Meanwhile, Givón (2001) and Dixon (2010) made a simplification by defining proposition as a clause that consists of a subject and a predicate. Proposition is also affected by clause-level operators that might modify the whole clause of the whole sentence (Pavey, 2010). Furthermore, Larson (1998) also suggested that a proposition often takes the form of a simple sentence in grammatical structure, even though this is not always the case. Larson (1998) also introduced the hierarchy of semantic structure to facilitate the

Language Literacy: Journal of Linguistics, Literature, and Language Teaching
Volume 6, Number 1, pp: 16-27, June 2022
e-ISSN: 2580-9962 | p-ISSN: 2580-8672
DOI: 10.30743/ll.v6i1.4678

understanding of a proposition. The following tables illustrate the hierarchy of semantic structure, which starts from the smallest unit of meaning:

Meaning component	Morphemes (roots and affixes)
Concept	Word
Complex concept (concept cluster)	Phrase
Proposition	Clause
Proposition cluster	Sentence
Semantic paragraph	Paragraf
Episode	Section
Episode cluster	Division
Semantic part	Part
Discourse	Text

Table 1. Hierarchy of *semantic structure* according to Larson (1998).

Based on the hierarchical semantic structure, Larson (1998) concluded that proposition is a semantic unit consisting of *concepts* (*thing, event,* and *attribute*. According to Givón (2001), there are three main categories of proposition, i.e., *State Proposition, Event Proposition,* and *Action Proposition.* However, the *Action Proposition* may also be part of the *Event Proposition* when it involves active *agent* as *participants* (Givón, 2001). Meanwhile, Larson (1998) categorizes proposition only into two large categories, i.e., *Event Proposition* and *State Proposition.* Within one proposition, one of the concepts becomes central and the remaining concepts are related to the central one through a system of *relations.* When the central concept is an Event concept, then the proposition will be categorized as *Event Proposition.* On the other hand, when the central concept is a *thing* or *attribute*, then Init is categorized as *State Proposition* (Larson, 1998). Larson (1998) also added that a state proposition consists of a *Topic* and a *Comment. Topic* is the *thing* or *attribute* being talked about, whereas *Comment* is what is being said about the *Topic* (Larson, 1998).

Therefore, in a simple clause, such as *"Budi loves Ani"* one can infer that there is one proposition embedded in that clause. This proposition consists of two concepts, i.e., *Budi* and *Ani* represent the concept of *thing* and *loves* represents the concept of *event.* This proposition can be encoded in different ways in the same language depending on the context, e.g., *"Ani is loved by Budi", or "The love Budi has for Ani…".* Either way of encoding, thus, can be deemed to contain the same proposition as *"Budi loves Ani".* If this proposition is to be translated into Indonesian language, then the way it is encoded can be *"Budi mencintai Ani", "Ani dicintai oleh Budi", "Cinta Budi untuk Ani…".*

2.2 Automatic evaluation vs manual evaluation of machine translation (MT)

Ever since the invention of the automatic translation system, many scholars from linguistics and computational domain have endeavoured to improve the system to produce a good translation with acceptable readership. Of course, the basis for making such improvement is the evaluation result of MT performance (Papineni et al., 2002). There are two main categories of MT evaluation, manual and automatic evaluation. One obvious distinction between the two categories is the tools being used and the baseline metrics.

One of the most common tool for conducting the automatic evaluation is BLEU (*BiLingual Evaluation Understudy)* (Papineni et al., 2002). This evaluation tool is perceived to be inexpensive, quick, applicable to any language, and significantly correlates with human evaluation (Papineni et al., 2002). The underlying idea of BLEU evaluation is that a "good"

machine translation is the one that is closer to human translation. Therefore, the method of conducting BLEU evaluation is by using the weighted average of variable length phrase matches against the reference translation or the *"gold standard"* translation (Koehn, 2010; Monz & Koehn, 2006). We can use this approach to generate an MT evaluation in terms of different word choices and word order.

Nevertheless, some studies have also been conducted in exploring the shortcomings of BLEU evaluation. Koehn & Monz (2006) conducted experimental research to compare the result of manual and automatic evaluation of MT between European languages. The automatic evaluation was done towards several state-of-the-art systems using BLEU, and the manual evaluation on fluency and adequacy was done by human assessor. This study reported that BLEU is not fully applicable for assessing the MT of all available systems (statistical MT, commercial rule-based system, etc.) (Monz & Koehn, 2006). This result confirms their general assumption that automatic evaluation is just an imperfect substitute for human assessment of translation quality. However, they have also found that human assessors also encountered difficulties when manually evaluating the MT output, particularly due to lack of correct reference translation and assessors' fluency in both TL and SL (Monz & Koehn, 2006).

Culy & Riehmann (2003) also conducted an experimental study that analysed the limits of *n-gram* translation evaluation metrics in BLEU and NIST tool. They examined the performance of professional human translations into German of two literary genres and compared the scores with the MT-output scores of the same source text. One surprising finding reported by this study is that non-fluent machine translation could score better than a completely fluent human translation. Therefore, they concluded that the low *n-gram* score does not necessarily indicate a poor translation, although a high *n-gram* score may indicate good translation. The high n-gram score was mainly given to translations that were fairly literal and preserved the word-to-word equivalence. Eventually, this study also answered the assumption that automatic evaluation is not applicable to measuring translation goodness. However, they can be used to measure document similarity (Culy & Riehemann, 2003).

In order to improve the evaluation method on MT, integrating scores of different measures must be done (Giménez & Màrquez, 2010). Therefore, Giménez & Màrquez (2010) proposed a new evaluation framework based on the concept of overlap among linguistic elements. They define linguistic elements as an abstract reference to any possible types of linguistic unit, structure, or relation between them. This framework is built upon an assumption that translation quality is represented by the number of lexical similarities to the reference translation and shares a similar syntactic and semantic structure. Eventually, this study recommended that measurement based on lexical, syntactic, and semantic similarity is more reliable than the evaluation, which only highlights lexical similarities.

A study on manual evaluation of MT output was also carried out by Popović (2020). She introduced a new method of manual assessment of MT through her study by marking all problematic parts of the text (words, phrases, sentences). This method is deemed the "mid-way" between overall assessment and error classification. However, this method can only address two quality criteria, i.e., comprehensibility and adequacy. Adequacy looks at how the meaning of the source text is conveyed in the translated text. In contrast, comprehensibility reflects readers' ability to understand the translated text without looking at the source text (Popović, 2020).

Language Literacy: Journal of Linguistics, Literature, and Language Teaching
Volume 6, Number 1, pp: 16-27, June 2022
e-ISSN: 2580-9962 | p-ISSN: 2580-8672
DOI: 10.30743/ll.v6i1.4678

2.3 Problems of non-equivalence

Even though the overall meaning of a translated proposition is not only determined by the word-to-word coherences between SL and TL, we still believe in the impact of word-level equivalence in the meaning delivery. A translated proposition might omit a word, paraphrase, or add an explanation as translation strategies that will affect the total word count but will not affect the overall meaning. However, the overall meaning of a proposition might be changed due to a mistranslation of a single word, phrase, or any lexical item above word and phrase level. In other words, there is a strong interrelation between word-level and above word-level equivalence and the overall meaning of a proposition.

When the target language has no direct equivalent for a word in the source text, this phenomenon is called non-equivalence (Baker, 2018). Baker (2018) identifies eleven types of non-equivalence: 1) *culture-specific concepts*; 2) *the source-language concept is not lexicalized in the target language*; 3) the SL word is semantically complex; 4) *the SL and TL make different distinctions in meaning*; 5) *the target language lacks a superordinate*; 6) *the TL lacks a specific term (hyponym)*; 7) *differences in physical or interpersonal perspective*; 8) *differences in expressive meaning*; 9) *differences in form*; 10) *differences in frequency and purpose of using specific forms*; 11) *the use of loan words in the source text.*

Different types of text or different language pairs may have different problems of non-equivalence. For instance, the problem of medical translation is more related to the fact that most medical terminology and medical collocations in particular are highly specialized even in the SL (Badziński, 2018) This claim is in line with Baker's non-equivalence category, i.e., the semantic complexity of the SL. On the other hand, the translation from Arabic into English faces different problems of non-equivalence, i.e., 1) culture-specific terms and concepts in the two languages, which include Islamic terms and concepts, Arabic customs, food, and social life; 2) Arabic terms which are not lexicalised in English; 3) semantic complexity of Arabic words; 4) different meaning distinction between Arabic and English; 5) Lack of hyponym in English; and 6) differences between Arabic and English in expressive meaning (Kashgary, 2011).

The categorization of non-equivalence coined by Baker (2018) is an invaluable help in identifying the underlying cause of problematic translation as part of the manual evaluation. By classifying the identified problems into a clear categorization, the result of MT evaluation can be used to improve the translation machines themselves. Meanwhile, the studies by Badziński (2018) and Kashgary (2011) provide a potential pattern and tendency of non-equivalence based on the specific text register and language pair.

3. Research Method

Ten academic-article abstracts (SL) on medical field with specific topic on palliative care were randomly selected as the data source for this study. These articles are written in English and can be accessed publicly through Sage Publishing website, one of the trusted sources of highly ranked journals. Medical writing is one of the registers, which contains highly technical terminology. Even though most professional healthcare professionals and medical scholars are capable of reading medical literature themselves, in some cases, the help of a translator may be required, especially when they do not have adequate proficiency in English (Daniele, 2019). The selection of medical writing is also based on the assumption that MT will never be used as a final and formal product. Instead, readers of medical writing or other technical topics will only use MT to obtain the gist of a text. Therefore, proposition-based evaluation is expected to be most applicable in translating text of this register.

The next step of data collection was translating the abstracts with Google Translate (GT). The automatic translation of all abstracts was done on the same date to avoid translation changes. The translated texts (TL) and the original text (SL) were split into sentences and transferred into Ms. Excel. The final step was conducting purposive sampling of the SL and TL sentences based on two categories. The first category is the sentence representing a full delivery of Event and State proposition from SL into TL. The second category is the sentence that contains a partial delivery of Event and State proposition.

The final data were analyzed with a qualitative approach and presented descriptively. In the first subsection of result and discussion, we present an exploratory elaboration of how proposition is conveyed into TL. In analyzing the surface structure, we adopt the theory introduced by Dixon (2010), whereas the deep structure embedded in the proposition is analysed by using Givón's (2001) semantic roles taxonomy and Larson's (1998) framework on proposition. The objective of this subsection is to reveal whether the translated sentences contain the same number of propositions as embedded in the SL or not. In the second subsection, we present the analysis of the causes that contribute to the successful delivery of constant proposition. In the third section, we present the analysis of the causes that contribute to the partial proposition delivery. The driving factors of success or failure in proposition delivery will be analysed by using the conceptual framework on *non-equivalence* coined by Baker (2014).

4. Results and Discussion

Based on a pragmatic-based approach, a translation process consists of two phases: interpreting the original SL text and constructing a target language text that conveys that interpretation (Farwell & Helmreich, 2003). Furthermore, the process of interpretation involves constructing a structure of propositions that convey the author's intended meaning (Farwell & Helmreich, 2003). Therefore, if an evaluation is to be conducted towards a translation output, it can be done by looking at the numbers of proposition(s) successfully conveyed in the translated text. The following subsection (4.1) describes how proposition is identified at sentence level and which type of proposition is embedded in the sample SL and TL texts.

4.1 When similar grammatical structures generate constant meaning in MT.

Given that a proposition, by most of linguistic scholars, is defined as a clause, then the way we identify it in a sentence is by revealing the clause, which constructed a sentence. Tree diagram is one feasible tool in making the breakdown of a sentence into clauses. However, when it comes to identifying the proposition, the categorization of each lexical item also involves tagging in accordance with its semantic roles. The details on how it is done are presented as follows:

Excerpt 1

ST: Constipation is one of the most common problems in patients receiving palliative care and can cause extreme suffering and discomfort.

MT: *Konstipasi adalah salah satu masalah paling umum pada pasien yang menerima perawatan paliatif dan dapat menyebabkan penderitaan dan ketidaknyamanan yang ekstrim.*

Language Literacy: Journal of Linguistics, Literature, and Language Teaching
Volume 6, Number 1, pp: 16-27, June 2022
e-ISSN: 2580-9962 | p-ISSN: 2580-8672
DOI: 10.30743/ll.v6i1.4678

According to its grammatical structure, Excerpt 1 represents a complex sentence consisting of two clauses connected with a coordinate conjunction: "*Constipation is one of the most commom problems in patients receiving palliative care"* as the main clause 1 and "*can cause extreme suffering and discomfort"* as the main clause 2*.* We can also identify that main clause 1 contains a relative clause as the object modifier: "*… receiving palliative care…".* We should bear in mind that this relative clause has omitted the relative pronoun *who* after the word *patient* as the *locative state* (Givón, 2001).

In terms of its propositional information, Excerpt 1 contains two propositions, i.e., the *State Proposition*, as embedded in main clause 1, and *Event Proposition,* as embedded in main clause 2. *Constipation* plays the role as the *patient of state* of the proposition and *"…one of the most common problems in patients receiving palliative care…"* is the *Attribute* of *Constipation*. This clause is classified as a *state proposition* since the *subject, which contains the meaning component of a Thing,* becomes the central concept of the proposition. Therefore, *Constipation* is the concept classified as the *topic*, i.e., whereas the rest of the clause represents *comment* since it is used to describe the topic*.*

Main clause 2 shared the same *subject* as the main clause, but it contains a separated predicate *(can cause)* with separated predicate argument *(extreme suffering and discomfort)*. A linkage marker position is filled by a coordinate conjunction *(and)* between the two clauses reflecting coordination relation. Despite sharing the same subject, the semantic role presented in main clause 1 and the semantic role "embedded" in main clause 2 are totally different. The word *Constipation* in main clause 1 takes the role of *patient of state,* whereas the subject of main clause 2 is an *agent.* This is because the omitted subject in main clause 2 is considered as the participant who acts deliberately to initate an event. Therefore, we can also infer that *can cause* is the concept of *Event* which becomes the central of the proposition and "*extreme suffering and discomfort"* is the *patient* of this proposition, or the *patient of change* to be more exact.

In excerpt 1, the MT contains the same number of propositions with the same sentence structure both grammatically and semantically. The MT also reflects a lot of lexical similarities and even sahres identical semantic roles with the ST, which results in almost the same word count between the two texts, i.e., 20 words in ST and 21 words in MT. From this data sample, we can infer that the preservation of grammatical structure can generate the same propositions, and thus fully deliver the meaning of ST, i.e., the truth/fact that constipation is a most common issue in palliative-care patients and can cause unbearable suffering and discomfort. The following excerpt shows another example of how the same grammatical structure between ST and MT positively affect meaning delivery in MT.

Excerpt 2

ST: Identification *of* the time point, where treatment *of* fatigue is no longer indicated is important to alleviate distress at the end of life.

MT: *Identifikasi titik waktu, di mana pengobatan kelelahan tidak lagi diindikasikan penting untuk mengurangi penderitaan di akhir kehidupan.*

Exerpt 2 of ST represents a simple sentence with a relative clause that functions as the object modifier. At the phrase level, we have found in Excerpt 2 that GT is already able to detect and transfer the grammatical function of the preposition "*of".* First, the preposition *"of"* from the noun phrase *"Identification of the time point"* expresses an association between *"Identification"* as the head noun and *"time point"* as the dependent noun or the

argument of the head noun. The same function of preposition *"of"* is also found in the noun phrase *"treatment of fatigue"* which also express *association* (Pavey, 2010). In this case, Google Translate seems to be able to detect the preposition "*of*" as an association marker, and then it omitted this lexical item in the Indonesian translation. Therefore, the translation found in the MT is *"Identifikasi titik waktu"* and *"pengobatan kelelahan".*

At phrase level, obviously, the form in MT has changed due to preposition omission. However, the meaning of the noun phrase remains constant. At the clause level, the grammatical structures of ST and MT are also similar. The noun phrase, which functions as the subject in the ST, i.e., *"Identification of time point",* is equivalently translated with the same grammatical function as the subject into *"Identifikasi titik waktu".* Even the relative clause *"…where treatment of fatige is no longer indicated…"* is translated with the same passive form in the MT. The predicate of the clause, *"…is important to alleviate distress at the end of life…"* is also translated into Indonesian with the same grammatical structure.

Based on the semantic roles constructing the proposition, we have found in excerpt 2 a positive correlation between grammatical structure and the delivery of propositional meaning. First, regardless the presence of a relative clause as the object modifier, there is only one proposition represented in excerpt 2. This proposition conveys the urgency to identify the time point within the palliative care where pharmacological treatment of fatigue is no longer needed because giving medication for fatigue at this time will only cause more pain to the patients. Similar to Excerpt 1, the subject in ST takes the semantic role as the *patient of state* and the object as the *attribute* of the *patient of state.* In addition, Excerpt 2 is categrized as *state proposition* since the central concept offered by the proposition is the *thing, which* is represented by noun phrase *"Identification of time point*". In this case, the *attributes,* or *comment,* describes *patient of state,* which also becomes the *topic* of the proposition.

Meanwhile, as mentioned above, the MT also shows the same grammatical structure, resulting in the same propositional information. Given the linearity of grammatical and semantic structure between ST and MT, thus this kind of translation is classified as literal translation. However, the result is the opposite of what Larson (1984) stated that a literal translation has little communication value. One factor that may enable literal translation as a nearly ideal translation is the genre of the text. Of course, we may find different result if the text used comes from literary genre. Nonetheless, so far, we can assume that literal translation can at least convey a constant meaning in a narrative text, particularly when the text contains fewer concepts, which carry secondary meaning.

4.2 When similar grammatical structure results in different meaning in MT

In the above subsection, we have presented the positive correlation between grammatical structure and the propositional meaning of an utterance in MT. Nevertheless, a natural translation often occurs with some changes at the surface structure, but still preserves the intended meaning of the original text. The following excerpt shows how symmetrical grammatical structure shifted the meaning in the translated proposition.

Excerpt 3

ST: This study aimed to explore **health professional, patient, family, and caregiver perceptions of palliative care**, availability of palliative care services to patients across South Dakota, and consistency and **quality of palliative care delivery.**

Language Literacy: Journal of Linguistics, Literature, and Language Teaching
Volume 6, Number 1, pp: 16-27, June 2022
e-ISSN: 2580-9962 | p-ISSN: 2580-8672
DOI: 10.30743/ll.v6i1.4678

MT: *Penelitian ini bertujuan untuk mengeksplorasi persepsi profesional kesehatan, pasien, keluarga, dan* ***pengasuh perawatan paliatif,*** *ketersediaan layanan perawatan paliatif untuk pasien di South Dakota, dan konsistensi* ***dan kualitas pemberian perawatan paliatif.***

In Excerpt 3, we have found a similar grammatical structure in the ST and MT. This is reflected from the similar grammatical function in each syntactic category in the ST and MT. For instance, both texts (ST and MT) have taken the passive form as the sentence construction, and each element has matched grammatical function. For instance, the *noun phrase* in ST, i.e., *"This study"* filled the grammatical function as subject, taking the semantic role as an agent. In the MT, the same semantic role and grammatical function is also taken by the literal translation of *This study*, i.e., *Penelitian ini.* Another evidence of grammatical structure similarities also shown in the modal verbs *"…aimed to explore..."* which is also translated into the same form, i.e., "*bertujuan untuk mengeksplorasi…".* This modal verb eventually represents the category of proposition: *Event Proposition.*

Interestingly, this time the similarity of form does not positively correlate with the meaning delivery in the MT. This finding is shown in the translation of "*…health professional, patient, family, and caregiver perceptions of palliative care…".* What is intended by the original author is that the study has the objective to explore the perceptions of all research subjects about palliative care and other topics (availability of palliative care, etc.). Unfortunately, the translation machine seemed to fail to detect the different function and meaning of the proposition *"of".*

We have seen that in excerpt 2, *of* becomes the proposition placed between a head noun and its argument. This construction eventually forms a noun phrase that expresses association relationship, thus being omitted in the MT. Google Translate uses this same algorithm in excerpt 3, resulting in meaning shift in MT. The meaning inferred in the first proposition of MT in excerpt 3 is that the purpose of the study is to explore the perception of the people involved in the palliative care. Meanwhile, the meaning that the study aimed to explore their perceptions about the palliative care itself is not conveyed. The translation also implies that palliative care is the argument noun of the health professional, patient, family, and caregiver, which has deviated from the meaning in ST. In this case, the preposition *"of"* should not have been omitted, and instead, should be translated in an equivalent form, such as *"mengenai"* or *"about"* in Indonesian. The following excerpt also shows another sample of meaning deviation in MT.

Excerpt 4

ST: Fatigue also plays a major role in palliative care for noncancer patients, with large percentages of patients with HIV, multiple sclerosis, chronic obstructive pulmonary disease or heart failure reporting fatigue.

MT: *Kelelahan juga memainkan peran utama dalam perawatan paliatif untuk pasien non-kanker, dengan persentase besar pasien dengan HIV, multiple sclerosis, penyakit paru obstruktif kronis atau gagal jantung melaporkan kelelahan.*

From the ST in excerpt 4, we can identify two propositions based on their overall meaning. The first proposition is that fatigue is also a huge problem in palliative care for

noncancer patiens. The second proposition is that most patients who have HIV, multiple sclerosis, chronic obstructive pulmonary disease, or heart failure also experience fatigue.

The first proposition is conveyed in the MT with symmetrical grammatical structure with ST but with loss of meaning due to the literal translation of the predicate *"plays a major role".* This verbal phrase reflects an *action* done by an *agent (fatigue)* in a *locative state (in palliative care for noncancer patients).* This semantic structure has eventually represented the *Event Proposition* since the central concept is the *action.* The translation replicates not only the grammatical structure but also the semantic structures. The Indonesian translation *"Kelelahan memainkan peran besar dalam perawatan paliatif untuk pasien non-kanker"* applied literalism approach, which could not convey the idea of fatigue as a problem for noncancer patients.

The ideal translation for this case has an implication to its semantic structure of the semantic roles of each participant and the proposition category. Given that the concept that must be highlighted is *Fatigue* and the severity of problem it brings to non-patient cancer, then a shift from *Event Proposition* into *State Proposition* should be considered, such as *"Rasa lelah merupakan permasalahan yang signifikan pada pasien non-kanker".*

4.3 Lexical inequivalence as the major cause of loss of meaning

Literal translation as the cause of shift/loss of meaning can be analyzed through the grammatical structure and the meaning component perspectives. In subsection 4.2, we have seen how identical grammatical structure resulted in inequivalent meaning of proposition. However, inequivalent translation at word level can also determine the success or failure of meaning transfer. We can still refer to Exerpt 3 to explain how inequivalent at word level has caused loss of meaning in MT.

The noun phrase ***caregivers*** actually refer to the individual who provides palliative care to patients. This individual can be someone from the patient's family, professional healthcare provider the family hires to provide care, or nurses at the hospitals. *Caregiver* is a concept used very specifically in the medical domain. Unfortunately, Bahasa Indonesia does not have the equivalent lexical item to accommodate all meaning components contained.

However, using the word *pengasuh* in the Indonesian translation is not ideal since the word is derived from the verb *"asuh"* which means to nurture. It refers to the process of caring and encouraging the growth of a being (someone or something). Meanwhile, a caregiver in the medical context, particularly palliative care, refers to a person providing treatment to patients in the recovery process or a terminal condition (end of life). Therefore, two strategies can be considered. The first one is by using the superordinate word *"perawat"* or back-translated as *nurse.* The second optional strategy is by explaining the meaning component within the clause or in a footnote. Another sample of how lexical inequivalent affects meaning transfer is also shown in the following excerpt.

Excerpt 5

ST: A **current** review of leadership opportunities **is provided**.

MT: *Sebuah tinjauan **saat ini** tentang peluang kepemimpinan **disediakan**.*

The back translation of MT in excerpt 5 is *"A review right now of leadership opportunities is provided".* Meanwhile, the meaning of the word *current* in the above context is the latest or the most up to date. People with good linguistic intuition probably

e-ISSN: 2580-9962 | p-ISSN: 2580-8672
DOI: 10.30743/ll.v6i1.4678

can guess the underlying meaning of the word *current* in that sentence. Still, this translation could be confusing for casual readers or laypeople.

Another problematic translation is the translation of passive form *"is provided"* which is translated into *disediakan* in the MT. First, the meaning of the proposition is that the study provides the latest review on leadership opportunity, in which *leadership opportunity* is an abstract idea of events and organizational management. In bahasa Indonesia, an abstract idea is not provided, but it is explained, described, or elaborated. Using the word *disediakan* will make the translation sounds unnatural to Indonesian readers and does not properly convey the meaning of the ST.

5. Conclusion

Automatic evaluation of translation products has long been and will always be an invaluable tool to the development of the automatic translation system. In this study, we are offering an approach that can complement the existing evaluation method. Our new approach highlights the proposition of an utterance as the vehicle of meaning. Using proposition as the unit for evaluation, we can see the quality of a translation not only from lexical similarities, which often leads to literal translation but also from the whole propositional meaning.

Using proposition-based evaluation, we can also identify the room for system improvement based on grammatical and semantic structure. From this exploratory study, we have found out that the similarity of grammatical and semantic structure between ST and MT positively correlates with the transfer of meaning in the target language, especially when the ST is a narrative text and contains fewer words with secondary meaning. However, the similarities of grammatical and semantic structure may also result in loss of meaning. Therefore, aiming for natural and equivalent translation may lead to the change of grammatical structure and/or even the change of proposition category.

Acknowledgements

We would like to express our ultimate gratitude to Dr. Sajarwa who introduced us to the realm of proposition and for encouraging us to see beyond what meets the eyes.

References

Badziński, A. (2018). *Collocations, Equivalence, and Untranslatability as Selected* Critical Aspects in Medical Translation. In *Towards Understanding Medical Translation and Interpreting*. Gdańsk University Press.

Baker, M. (2018). *In Other Words*. Routledge.

Chomsky, N. (2019). Studies on Semantics in Generative Grammar. In *Studies on Semantics in Generative Grammar*. De Gruyter Mouton. https://doi.org/10.1515/9783110867589

Culy, C., & Riehemann, S. Z. (2003). The Limits of NGram Translation Evaluation Metrics. *Proceedings of Machine Translation Summit IX*, 71–78.

Daniele, F. (2019). Performance of an automatic translator in translating medical abstracts. *Heliyon*, *5*(10), e02687. https://doi.org/10.1016/j.heliyon.2019.e02687

DeFrancesco, C., & Perkins, K. (2012). An Analysis of the Proposition Density, Sentence and Clause Types, and Non-Finite Verbal Usage in Two College Textbooks. *Proceedings of the 11th Annual College of Education & GSN Research Conference*, 6. http://education.fiu.edu/research_conference/

Dixon, R. M. W. (2010). *Basic linguistic theory* (Vol. 2). Oxford University Press.

Farwell, D., & Helmreich, S. (2003, September 23). Pragmatics-based translation and MT evaluation. *Workshop on Systemizing MT Evaluation*. MTSummit 2003, New Orleans, USA. https://aclanthology.org/2003.mtsummit-eval.3

Giménez, J. & Màrquez, L. (2010). Linguistic measures for automatic machine translation evaluation. *Machine Translation*, *24*(3–4), 209–240. https://doi.org/10.1007/s10590-011-9088-7

Givón, T. (2001). *Syntax: An introduction. Vol. 1: ...* (Rev. ed). Benjamins.

Kashgary, A. D. (2011). The paradox of translating the untranslatable: Equivalence vs. non-equivalence in translating from Arabic into English. *Journal of King Saud University - Languages and Translation*, *23*(1), 47–57. https://doi.org/10.1016/j.jksult.2010.03.001

Kemp, G. (2018). *What Is This Thing Called Philosophy of Language?* (2nd ed.). Routledge.

Koehn, P. (2010). Statistical Machine Translation. Cambridge University Press. https://www.cambridge.org/id/academic/subjects/computer-science/artificial-intelligence-and-natural-language-processing/statistical-machine-translation?format=HB&isbn=9780521874151

Larson, M. (1998). *Meaning-based Translation. A Guide to Cross-Language Equivalence* (2nd ed.). University Press of America.

Monz, C., & Koehn, P. (2006). Manual and automatic evaluation of machine translation between European languages. *Proceedings of the Workshop on Statistical Machine Translation*, 102–121.

Papineni, K. et al. (2002). Bleu: A Method for Automatic Evaluation of Machine Translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311–318. https://doi.org/10.3115/1073083.1073135

Pavey, E. L. (2010). *The structure of language: An introduction to grammatical analysis*. Cambridge University Press. https://doi.org/10.1017/CBO9780511777929

Popović, M. (2020). Informative Manual Evaluation of Machine Translation Output. *Proceedings of the 28th International Conference on Computational Linguistics*, 5059–5069. https://doi.org/10.18653/v1/2020.coling-main.444

Wu, Y., & Xu, R. (2011). The Application of Chomsky's Syntactic Theory in Translation Study. *Journal of Language Teaching and Research*, *2*(2), 396–399. https://doi.org/10.4304/jltr.2.2.396-399