AJELP: The Asian Journal of English Language & Pedagogy
ISSN 2289-8689 / e-ISSN 2289-8697, Vol 9 No.2 (2021),109-127

Google Translate vs. DeepL: A quantitative evaluation of close-language pair translation (French to English)

Ahmad Yulianto

Faculty of Languages and Arts
Universitas Negeri Semarang (UNNES), Indonesia

Rina Supriatnaningsih

Faculty of Languages and Arts
Universitas Negeri Semarang (UNNES), Indonesia

e-mail: ay@mail.unnes.ac.id , rinasupriatnaningsih@mail.unnes.ac.id

Published: 20 December 2021

To cite this article (APA): Yulianto, A., & Supriatnaningsih, R. (2021). Google Translate vs. DeepL: A quantitative evaluation of close-language pair translation (French to English). *AJELP: Asian Journal of English Language and Pedagogy*, *9*(2), 109-127. https://doi.org/10.37134/ajelp.vol9.2.9.2021

To link to this article: https://doi.org/10.37134/ajelp.vol9.2.9.2021

Abstract: Machine translation has improved in quality and worked best when applied to language pair of the same language family. This research was aimed to assess the quality of Google Translate and DeepL in terms of accuracy and readability. French to English translation data of *En attendant Godot* playscript by GT and DeepL were evaluated. The English Original version (EO) of the text served as reference. Two quantitative methods were employed i.e., manual with SAE J2450 translation metric and automatic assessment with Coh Metrix tool. The result of manual assessment shows that GT and DeepL outputs passed the grade, scoring 84 and 99.04 respectively. Referring to CdT Rubric, a translation is good when it has 80 - 99 points. In Coh-Metrix result GT and DeepL scores varied. Statistical analysis with ANOVA shows that GT and DeepL are not significantly different from EO. EO mean score is 99.69, GT is 100.4 and DeepL is 100.78. In conclusion, DeepL scores higher in manual assessment, indicative of its accuracy while GT and DeepL are more or less the same in Coh-Metrix assessment. In terms of readability, DeepL offers better reading ease as proved by Flesch Reading Ease, Flesch-Kincaid Grade Level and Coh Metrix Readability formulas, all in favor of DeepL. Despite this statistical result, there are many things that GT and DeepL need to improve like world knowledge and ability to decipher lexical and structural ambiguities.

Keywords: assessment, evaluation, human text, machine translation, metric

INTRODUCTION

Need for translation has been increasingly growing in recent years. A huge amount of data on the internet demands for fast, accurate and reliable conversion from one language to another. Bell (1991) contended that the main objective of translation is to transform an original text in one language into its equivalent in another language so as to convey the meaning, its formal features and functional roles of the original text. So, when we translate from source language (SL) to target language (TL) some issues may arise such as culture-specific items, lexical equivalent, sentence structure, etc. Nevertheless, for languages that are close to one another, some of these problems are probably absent.

Each work of translation has its own difficulties and challenges. Crystal (1991: 346) stated that translation is a process where the meaning and expression in the source language is adjusted with that in the target language. Torop (2002) has maintained that translation is rooted in the sociocultural language of a particular context since it is a process of converting ideas expressed from one language into another. Translation is the reproduction of the closest natural equivalence of the source language in a target language both in terms of meaning and style (Nida and Taber, 1969: 208). So, it is not easy to determine what a good translation is since even two professional translators may produce different outputs from the same source text. Human translation (hereinafter HT) is frequently time-consuming and costly. In this respect machine translation (hereinafter MT) has apparently offered a possible solution to the painstakingly process of human translation.

Nowadays people are becoming more dependent on machine translation with Google Translate (hereinafter GT) as the most-widely used platform. Over 200 million users and a billion translations per day were reported in 2013 (Shankland, 2013 cited in Li, Graesser and Cai, 2014). Some experts claim that MT output including GT is almost human-like in the sense that it gets closer to version of average human translators (Wu et al., 2016). Some others argued, saying that MT serves only for limited purposes (Puchała-Ladzińska 2016: 95). It cannot reach the quality and naturalness of human work and human correction is still needed in most cases. Puchała-Ladzińska added that MT has not yet attained the quality standard of professional translators. It is generally believed that MT tends to work best on certain types of text like technical, legal, commercial, manual, instruction, and the like (Štefčík, 2015: 140).

BACKGROUND OF THE STUDY

This study was aimed to assess the outputs of MT namely GT and DeepL in terms of accuracy and readability. Literary text was chosen to find out whether the claim over MT is correct. Translating literary work is believed as a difficult process and is deemed as an artistic and creative practice. Thus, it is necessary for the translators to make a kind of contemplation over the esthetics of the translated text to keep its style more or less the same. In addition to excellent language abilities, they are expected to possess artistic tastes as well (Fowler & Hodges, 2011). All the above-mentioned aspects seem to be human's properties.

Assessing translation is a daunting yet essential task not only for translation quality improvement but also research in translation. Traditionally, we depend on our general impression on a text in order to measure its quality. Translation assessment method itself dates back to 1966 with the main emphasis lies on intelligibility and fidelity (Carroll, 1966). To be intelligible means that translation should read as normal as possible and be easily understood.

In fact, there is no common ground of defining translation quality both from practical and theoretical point of view. To many scholars, quality in translation is a subjective concept (Horguelin and Brunette 1998; Larose 1998; Parra 2005). However, experts have agreed on what measures should be taken when creating a reliable assessment. There are at least three steps to be taken: firstly, quality must be defined. A quality translation is when it fits its purpose (Nord 1997; O'Brien 2012). Secondly, the methodology must be set. For that reason, special attention has to be given to assessment methods that enable measurement. And thirdly, the assessment should be carried out in accordance with the definition of quality as applied to the text and to the assessment methodology chosen (Martinez, 2014: 73).

In recent years there are two trends in assessment methodology for translation. The first trend scrutinizes the linguistic features of the translated texts at sentence level by means of error-based translation evaluation system as the standard procedure for quantifying quality (Secâra 2005), while the other focuses on macrostructure relations of the text as a unit. Waddington (2000) named the first type as quantitative-centered (bottom-up) system and the second as the qualitative centered (top-down) system. According to Williams (2004) the first type refers to error counting whereas the second is holistic systems.

Assessment of MT is also important for it may help to determine whether the existing MT systems generate acceptable translations or not. It also helps to find out which parts of MT should be improved for better performance. In the early 1990's human subjective judgments were used to score the semantic accuracy and fluency of MT outputs against one or more professionally produced human reference translations. However, human judgment is likely to fall into subjectivity and less measurable. As information technology and artificial intelligence have progressed rapidly, more and more instruments for computational MT evaluation are created. It is becoming common now to use tools to help us assess MT translation quality.

SAE J2450 Metric

Originally designed for the automotive industry, SAE J2450 metric has gained popularity and attracted the attention of many scholars in translation studies (Sun, 2015: 43). This metric provides a translation error scoring system to measure the quality of translation regardless what source language or the target language is, and how the translation is performed (i.e., HT or MT). It has seven primary error categories and two error classifications (namely, minor or serious). These error categories include wrong term (WT), syntactic error (SE), omission and addition (OM/AD), word structure (WS), misspelling (SP), punctuation error (PE), and miscellaneous error (ME) (SAE 2450, 2001: 2).

Term is defined as any single word, multi-word phrase, abbreviation, acronym, number or numerals, or proper name. A wrong term (WT) is defined to be any target language term that: (1) infringes term glossary; (2) does not conform with conventional or professional usage; (3) does not conform with other translations of the source language term; (4) denotes a concept which is different from the concept in the source language. Syntactic Error (SE) refers to wrong parts of speech, phrase structure, or order of words. Omission (OM) and addition (AD) occur when a block of text in the source text has no counterpart in the target language so that it is not translated in TL or a word/term is added. Word Structure (WS) refers to an incorrect target language word or an incorrect form of a target language term, such as the wrong use of upper- and lower-case letters, gender, numbers, tense, and prefixes, suffixes and infixes. Misspelling (SP) error includes infringement of the spelling, of the accepted norms for spelling in the target language, and infringement of the appropriate writing system. Punctuation errors are determined according to the punctuation rules of the target language. Miscellaneous Error (ME) covers what is not clearly classified under the previous categories like literal translations of idiomatic expressions which may be linguistically accurate but culturally inappropriate.

In addition to these categories, this metric classifies errors as serious or minor. A serious error is one which produces the wrong meaning, leading to confusion for the user and creating a risk of doing the wrong thing. A minor error will only lead to slight confusion or no confusion at all. Each error is then assigned with different weighting. The most serious mistake in WT is weighted 5 points and the least serious mistake is weighted 2 points. The most serious error in SE, OM, and WS is scored 4 points while the least serious one is scored 2 points. SP and ME share the same weighting of 3 points for the most serious error and 1 for the most minor mistake. Scoring in PE is set 2 points for serious error and 1 point for minor error. To calculate the total score, the weight (W) for determined each category is multiplied by the frequency (F) or how many times this type of error occurs. Finally, an overall score is tallied according to the weighted scores in all seven categories. A high score indicates that the translation is of low quality (SAE J2450, 2001: 3).

Coh-Metrix

Coh-Metrix is a computational tool that generates linguistic indices and discourse representations of a text. It contains 108 indices that are categorized into eleven groups: (1) Descriptive, (2) Text Easability Principal Component Scores, (3) Referential Cohesion, (4) LSA, (5) Lexical Diversity, (6) Connectives, (7) Situation Model, (8) Syntactic Complexity, (9) Syntactic Pattern Density, (10) Word Information, and (11) Readability. The output can be used in various ways to study the cohesion of the text and the coherence of mental representation of the text (Graesser, McNamara, & Louwerse, 2003). Coh-Metrix integrates a wide variety of modules used in computational linguistics (Graesser et al 2004: 194). Coh-Metrix pays attention to a wide range of linguistic features that affect comprehension such as cohesion, world knowledge and discourse characteristics (Graesser et al 2014: 194).

Accuracy in translation refers to the correctness of the meaning or message that is transmitted in translation (Arnold et al, 1994:162). Rahimi (2004) stated that accuracy refers to the appropriate and detailed description of the source text and its precise transfer to target text. Rahimi added that a translation should be deemed inaccurate if it neglects some pieces of information or adds what is not found in the source text. Khomeijani (2005) believes that accuracy refers to how precise the translator translates a text.

Larson (1984:482) proposed four indicators of inaccuracy in translation, namely omission, addition, different/wrong meaning, and zero meaning. (1) Omission is characterized by the absence of one or more items that otherwise must appear in the translated text. (2) Addition is marked with the presence of one or more items in the target text for getting the meaning across. (3) Different/wrong meaning occurs either in the analysis of the source text or in the transfer process. (4) Zero meaning is characterized with the use of form that does not convey any message at all.

In translation, readability is concerned with the understanding of source text and target text. According to Richards et. al. (1997: 62), readability refers to how easy written materials can be read and understood. There are some factors that determine readability of a text. Richards et. al. (1997: 63) confirms that readability depends on the average length of sentences, the number of new words and language grammatical complexity. A text which comprises unusual words will be difficult to understand by the readers. So, a complex sentence will be more difficult to understand than the simple one.

Considering the background, the research questions were the following: (a) how accurate the outputs of GT and DeepL with respect to EO? (b) which machine translation Google Translate or DeepL is better in translating literary work of close language pair, French - English? (c) which output offers reading ease GT or DeepL?

METHOD

Data

The data in this study consist of 4 corpora. The first is the source text (hereinafter ST) in French, taken from Samuel Beckett's *En attendant Godot* book, *Edition de minuit*, published in Paris, 1952. The second corpus is the English original version (hereinafter EO) of the same text taken from *Waiting for Godot* book. This EO was written by Samuel Beckett 2 years after he wrote it in French. It was published by New York's Grove Press in 1954. In this study EO serves as a reference which the evaluation of GT's and DeepL's output must refers to. The other corpora include 2 MT outputs generated by GT and DeepL on July 31, 2021.

To analyze the data, two different methods were used. The first one was manual judgment done by a human rater while the second one was computational assessment conducted by a tool. The result of this manual judgment on accuracy was then confirmed with the automatic evaluation. Readability level was fully measured by the tool. As mostly occurring in MT assessment, proximity to human translations is deemed highly essential (Finch, Hwang and Sumita 2005: 17).

The instruments employed in this study consist of SAE J2450 Translation Quality Metric and Coh-Metrix 3.0. tool. Seven primary error categories SAE J2450 namely wrong term (WT), syntactic error (SE), omission and addition (OM/AD), word structure (WS), misspelling (SP), punctuation error (PE), and miscellaneous error (ME) were used.

Twelve indices of Coh-Metrix namely paragraph count, mean of paragraph length, sentence count, mean of sentence length, word count, mean of word length, noun incidence, verb incidence, adjective incidence, adverb incidence, type-token ratio, and Flesch Reading Ease were used. The first 11 indices enable us to predict accuracy while the last three indices deal with readability.

Procedures

For this study is a combination of manual and computational assessments, then several steps must be followed strictly to make sure that the analysis is performed as set out and yields the result as expected. First of all, four corpora were juxtaposed and presented in a table: the source text (in French), EO, and two English translation outputs from GT and DeepL. EO served as reference in judging manually GT and DeepL outputs. These raw data from MT outputs were then observed to identify any translation errors. These translation errors were marked by means of SAE J2450 Translation Quality Metric. After that, types and number of errors were tabulated to help count the error frequency and determine the weighting. Finally, an overall score was tallied according to the weighted scores in all seven categories. A high error score shows low quality translation. Based on this final score, GT and DeepL outputs were determined; whether they are of poor, medium or high qualities.

Afterwards, computational assessment was performed by using Coh-Metrix 3.0. Since these four corpora were from different sources and the analysis involved computation, special attention had to be paid to the texts' characteristics to make sure that they meet the requirement set by the tool. For that purpose, these 4 corpora were converted to *.txt* format. After that, they were inserted into Coh-Metrix tool for indices identification. The result table was then saved and analyzed. Not all the data yielded were used. Only those deemed pertinent to the goal of this study were taken. Finally, the score of each index were analyzed and interpreted into meaningful representations.

FINDINGS AND DISCUSSION

SAE J2450 Metric Data Analysis

As stated earlier, the first analysis was done by means of SAE J2450 Metric. Tabulated data is presented in the table below.

Table 1: Translation Evaluation on SAEJ 2450 Metric

	GT Error		**DeepL Error**		**Total Score (No of Error x Weight)**	
	Serious	**Minor**	**Serious**	**Minor**	**GT**	**DeepL**
WT	10	5	3	1	61	17
SE	1	2	1	1	8	6
OM/ AD	-	3	-	2	6	4
WS	-	1	-	1	2	2
SP	-	-	-	-	-	-
PE	-	-	-	-	-	-
ME	-	-	-	-	-	-
Total	11	11	4	5	**77**	**29**

As shown in the table, there are 15 GT wrong term errors of which 10 are serious and 5 are minors. The most conspicuous and unacceptable mistake is probably the substitution of proper name Estragon, a character in the play, into Tarragon which recurs several times. No clear explanation can be made for this substitution since proper names are usually kept intact in translation. The second mistake considered serious is the translation of French phrase "*même jeu*". It is true that literally it means "*same game*" as found in the GT. Yet, contextually that phrase means "*as usual, as always*", etc. as it is confirmed in EO in which it was translated "*as before*". The third mistake is the translation of adverb "*tout à l'heure* or *à tout à l'heure*" in its complete form. Literally *tout à l'heure* means "*just now* or a *while ago*" while *à tout à l'heure* means "*later, shortly* or *see you later*". In French their uses are interchangeably. GT was unaware of this and mistranslated it to "*All on time, on time*" which deviates in meaning from its original message. The next serious mistake GT made is keeping the word "*Monsieur*" in the text instead of finding its correct equivalent in English. *Monsieur* has various meanings and usages. It can be used to address a man like "*Monsieur*" which is similar to title Mister or Sir in English. On its own, it also means gentleman, master or lord (as a form of honorific title). The next mistake is translation of "*depuis le temps*" into since time. Indeed, on word-by-word translation "*depuis*" means "*since*" and "*le temps*" means "*the time*" but "*depuis le temps*" is an adverb of time which should actually translate into "*all the time*". The last major mistake GT made is the translation of "*Et après*?" into "*and after*?". This is a kind of meaningless translation which contextually doesn't make any sense at all. The correct translation is "*and then?*" or "*then what?*"

Unlike GT, DeepL is a way better. In Wrong Term category it only made 4 mistakes of which 3 are considered serious. The first mistake it made is just the same as GT namely the translation of French phrase "*même jeu*" translated into *same game*. "As usual or as before" is more appropriate. Translation of "*Tout à l'heure, tout à l'heure* made DeepL second mistake. That French expression was translated into "*just now, just now*" which implies that the action has already been done. In fact, it is just on the contrary. The character in the play named

Estragon says "*Tout à l'heure, tout à l'heure*" to tell his interlocuter to do the action later or not now. The last mistake DeepL made is the translation of the word "*monsieur*" into "*sir*" in the following sentence "*Peut-on savoir où monsieur a passé la nuit*?" (*Can we know where sir* spent the night?*). The word "*monsieur*" has various meanings and usages. It is used to address a man like in "*Monsieur Hercule Poirot"* which is similar to "*Mr. Hercule Poirot*" in English.

In terms of Syntactic Error GT also made more mistakes than DeepL with the score 3 against 2. GT made 3 errors while DeepL 2 errors. The GT first translation mistake is on "*Lève-toi que je t'embrasse*" translated into "*Get up and I kiss you*." Instead, it should be "*Get up so that I can kiss you*." GT failed to comprehend the message conveyed in this sentence including relative pronoun "*que"* which means "*so that*" in this case. The second mistake of GT is the translation of French adverb "*avec irritation*". Most English adverbs end with suffix *-ly* while in French only some adverbs end with suffix *-ment*. Other adverbs of manner have to be expressed by using the word "*avec*" (literally: with) or "*de manière*" (literally: in a way or in the manner).

So, the correct translation of "*avec irritation*" (Fr) is "*irritably*" (Eng) since it refers more to "*a feeling of being disturbed* or *annoyed*" rather than "*having physical skin inflammation*." The third error GT made is not serious for it only involves the modification of the original negative interrogative sentence without changing its overall message: "*Ça ne t'est jamais arrivé à toi?*" (*Has it never happened to you?*) which was translated into "*Has it ever happened to you?*" DeepL made less errors in this category, one serious and one minor. French phrase "*Route à la campagne*" preferably translated into English noun adjunct "country road" was translated by DeepL as "*A road in the country*." The second structural error of DeepL translation is just the same as what occurred in GT, that is the translation of "*avec irritation*".

In terms of Addition & Omission both GT and DeepL do not much deviate much from the source text. GT made 2 additions while DeepL only made 1 addition. The first addition is found in the phrase "*Route à la campagne*" meaning "*country road*" but was translated by GT and DeepL into "*A country road* and *a road in the country*" respectively. The second addition is found in the works of GT and DeepL and is concerned with *Pronom Tonique* or French Stressed Pronouns. It is common in French to say: "*Moi, je ne sais pas.*" meaning "*I don't know*" with *Pronom Tonique* – "*moi"* put in front of or at the end of the sentence.

This French pronoun is sometimes translatable into objective pronoun, "*just*" or not need translating at all. The word "*moi*" literally means "me". However, the sentence "*Moi, je ne sais pas*." can be translated into "*I just don't know*" with "just" serves as the equivalent of the French *Pronom Tonique* "*moi*" in order to emphasize that the subject "*I really don't know*".

For word structure error category GT and DeepL made 1 error respectively that occurred on the same word "*tries*". It should be "*trying*" as equivalent of French word "*essaie*" as shown in this sentence "*Estragon, assis sur une pierre, essaie d'enlever sa chaussure*." (*Estragon, sitting on a stone, was trying to take off his shoes*). No spelling, punctuation or miscellaneous errors was found either in GT or DeepL.

Coh-Metrix 3.0 Data Analysis

In this automatic assessment EO data were calculated and presented in the table as well as it was required by the tool as comparison. EO was not judged but served as a reference or comparison to GT and DeepL. See the table below.

Table2: Coh-Metrix 3.0 Analysis

No	Indices	EO (SD)	GT (SD)	DeepL (SD)
1	Par. Count	35	40	39
2	Mean of Par. Length	2.457 (1.721)	2.100 (1.336)	2.128 (1.321)
3	Sent. Count	86	84	83
4	Mean of Sent. Length	5.384 (4.258)	5.619 (3.997)	5.771 (3.887)
5	Word count	463	472	479
6	Mean of Word length	4.225 (2.264)	4.258 (2.264)	4.225 (2.309)
7	Noun Incidence	188	201	192
8	Verb Incidence	158	161	157
9	Adjective Incidence	52	47	48
10	Adverb Incidence	102	87	98
11	Type-token ratio	0.467	0.464	0.448
12	Flesch Reading Ease	82.761	82.861	83.891
13	Flesch-Kincaid Grade Level	3.053	3.098	2.992
14	Coh-Metrix L2 Readability	19.037	21.085	21.432

***SD** is standard deviation.*

As can be seen in the Table 2, there are 14 Coh-Metrix indices employed in this analysis. These indices are considered relevant to accuracy and readability. Indices No. 1 – 11 strictly deal with accuracy measurement whilst 12 – 14 are concerned with readability. The first index namely *Paragraph Count* refers to the number of paragraphs in the text. The table shows that the number of paragraphs varies in EO, GT and DeepL. EO has the least number of paragraphs containing only 35, followed by DeepL consisting of 39 and GT comprising 40 paragraphs. In this concern it is interesting to see how each text differs in paragraph length as indicated in the EO Mean of Paragraph Length, i.e., 2.457. GT Mean of Paragraph Length is 2.100 whereas DeepL's Mean is 2.128. Although all corpora have 2 – 3 sentences, it clear that EO is relatively longer.

Standard deviation (hereinafter SD) explains further how one text is different from the others in terms of length. A large standard deviation indicates that the text has large variation in terms of paragraph length. We learned from the text that both GT and DeepL turned out to have smaller standard deviation of 1.336 and 1.321 respectively, meaning that their texts are less various in paragraph length. In other words, the paragraphs that make up the texts tend to be similar in length. On the contrary, the reference EO has the largest standard deviation which is 1.721. It means that the paragraphs in this text can be either so short or so long or greater in variation.

The second index, *Sentences Count*, represents the number of sentences in the text. DeepL calculated 83, GT 84 and EO 86 sentences. It is comprehensible if EO has more sentences for human translators oftentimes break down longer sentences into shorter ones for the sake of clearer meaning. While GT and DeepL have fewer sentences, but their Mean of Sentence Length are larger than EO with 5.771 and 5.619 each, compared to 5.384 of EO. It suggests that each sentence in these 2 MT outputs contain 5 – 6 words. To be more specific, GT output consists of 5.771 words whereas DeepL output comprises 5.619 words on average.

Concerning SD, it can be stated that a large standard deviation indicates that the text has large variation in terms of its sentence length, such that it may have some very short and some very long sentences. DeepL has the smallest SD of 3.887 which indicates that the sentences in DeepL output are relatively similar in length, or not too various. GT comes in second with SD of 3.887. So, its sentences are more various in terms of length; some are long

while some others are short. EO has the biggest SD of 4.258, indicating that its text is much greater in variation; some sentences are so short while some others are quite long. Above of all, it should be noted that long sentences are usually more complex and therefore become more difficult to be understood by the readers. Here are the examples of the shortest sentences and the longest ones.

Table 3: Sentence length comparison

Sentences	Source Text (ST)	English Original (EO)	Google Translate (GT)	DeepL
Short	Tu crois ?	Am I?	You think so?	You think so?
Medium	Je commence à le croire.	I'm beginning to come round to that opinion.	I'm starting to believe it.	I'm beginning to believe it.
Long	Il s'arrête, à bout de forces, se repose en haletant, recommence.	He gives up, exhausted, rests, tries again.	He stops, at the end of his strength, rests panting, starts again.	He stops, out of strength, rests panting, and starts again.

Table 3 contains example of sentences taken from the 4 corpora which are different in length. It should be kept in mind that EO in this context is not human translation *an sich*, rather it is the English version of the same text produced by the same author of the French original text. As presented in the table above the shortest sentence in both ST and HT which happen to be interrogative sentence "*Tu crois?*" and "*Am I?*" Each consists of 2 words. '*Tu*' as subjective pronoun '*crois*' as verb/predicate. It goes the same way with "*Am I*?". '*Am*' is to be or verb while '*I*' is subjective pronoun. To convey the same message GT and DeepL use 3 words where they use the same sentences namely, "*You think so?*". At glance, GT and DeepL seem to have better translated the sentence than EO for the words '*tu*' and '*crois*' literally mean '*you*' and '*think*' or '*believe*' respectively. Yet, it cannot be separated from the context and previous sentence which reads "*Alors, te revoilà, toi*." or "*So, there you are again*". The sentence "*Tu crois?*" is actually a response to the previous remark. For this sentence, GT and DeepL succeeded in making appropriate translation by relying on the structural and lexical patterns. EO, as a product of human creativity, uses a different expression which is contextually correct and acceptable although structurally different.

For the medium-long sentence, GT and DeepL sentences have exactly the same number of words as ST, that is 5 words. "*Je commence à le croire.*" translates into "*I'm starting to believe it*." in GT output and "*I'm beginning to believe it*." in DeepL output. The difference of the 2 MT lies in the different choice of French word '*commence*' equivalent which is translatable into either '*starting*' or '*beginning*'. DeepL output is a bit better for the word '*beginning*' is preferable in this case since the message suggests that it happens naturally. However, statistically these 2 MT are similar.

For long sentence category GT and DeepL are more or less the same in sentence length. GT output is *He/stops, /at the end of/his/strength, /rests/panting, /starts/again* whereas DeepL output is *He/stops, /out of/strength, /rests/panting, /and/starts/again*. These 2 sentences consist of about 9 words but interestingly a little different. To make it clearer, let's compare it with the source text *Il/s'arrête, /à bout de/forces, /se repose/en haletant, /recommence* which comprises 7 words. GT and DeepL generated the same translation for the first 2 words but did it differently for the expression '*à bout de force*'. GT prefers precisely literal translation '*at the end of his strength'* whereas DeepL favors a more contextual translation, that is, '*out of strength*'. The reason why the English translation has more words is partly due to its verb characteristics which differs from French. English is rich in phrasal verbs such as *get up, get over, go by*, etc. while French is not. Instead of using combination of root plus preposition like in the word *start again* (which is computationally counted as 2 words/tokens), French has the word *recommencer* (which is computationally counted as 1 word/token) to express the same idea.

The next index is *Word Count* which indicates the number of words. EO has the least number of words, that is 463; less than GT with 472 and DeepL with 479. It is surprising though to see that EO has less words than the 2 MT since human beings usually tend to write more and their text consequently becomes flowery or wordy. Relative to Mean of Word Length, GT scored 4.258, meaning that the words in GT output are of 4 – 5 syllables; with its precise length average is 4.258. DeepL and EO scores are the same, namely 4.225. So, each word in these 2 texts is of 4.225 syllables on average. Implication is that shorter words are usually easier to read and understand. A large SD in *Word Count* indicates that the text has large variation in terms of the lengths of its words, such that it may have both short and long words. Table 3 shows that DeepL has the largest SD namely 2.309, meaning that words used in DeepL output are more various and relatively different in length. Some words are probably of few syllables (one or two syllables) whereas some others longer words (three or more syllable-words). On the other hand, GT and EO shares the same SD score, that is, 2.264. It suggests that they consist of words which are more or less the same in length.

Take a look at this example which shows how GT is different from DeepL in terms of word length. To express idea of French word '*se recueille*' (3 syllables) GT uses the verb '*recollects himself*' (5 syllables) while DeepL prefers the verb '*collects himself*' (4 syllables). '*Se recueillir*' (Fr) is a reflexive verb where '*se*' refers to the agent or the person which in English may translates into myself, yourself, himself, herself, ourselves or themselves depending on the person. The second example if the translation of verb '*savoir*' (2 syllables) into '*find out*' (2 syllables) by GT and '*know*' (1 syllable) by DeepL. In this case GT seems to prefer contextual translation while DeepL favors literal translation. Regardless the meaning, this difference shows the reason why GT has bigger mean of word length than DeepL.

Noun incidence indicates the frequency of noun occurrence in the text. Noun appeared 188 times in EO, 201 times in GT and 192 times in DeepL. Verb incidence indicates how many times verb shows up in the text. EO contained 158 occurrences of verb, GT noted 161 occurrences and DeepL had 157. Adjective incidence displays the appearance of adjective in a particular text while adverb incidence indicates the frequency of adverb. It is interesting to reveal that EO used adjective 52 times, much more frequent than GT with 47 times and DeepL which was 48 times. It is highly likely that human being is more creative than machine and therefore like to use adjective to describe things. In adverb incidence category adverb appeared 102 times in EO, 87 times in GT and 98 times in DeepL.

Type-token ratios of these 3 translations are quite similar with EO ratio of 0.467, GT of 0.464, and DeepL of 0.448. Type-token ratio is the number of unique words (called types) divided by the number of tokens of these words (Templin, 1957). Each unique word in a text is considered a word type. Each instance of a particular word is a token. For example, if the word dog appears in the text 7 times, its type value is 1, whereas its token value is 7. When the type-token ratio approaches 1, each word occurs only once in the text; comprehension should be comparatively difficult because many unique words need to be decoded and integrated with the discourse context. As the type-token ratio decreases, words are repeated many times in the text, which should increase the ease and speed of text processing. In other words, the text is easier to understand. TTR scores are most valuable when texts of similar lengths are compared. So, DeepL suggests more ease in text comprehension while EO is relatively more difficult to understand.

One thing that needs further investigation is how these 3 corpora show different occurrences in part of speech. We are not assessing EO since it is beyond the scope of this study. Either GT or DeepL, or may be even both, identified wrongly parts of speech in the source text. Or perhaps some parts of speech in the source text cannot be found its equivalents

in the source text so that MT decided to use different parts of speech to convey the message. To be considered accurate target text should not be much different from source text.

This is also confirmed with the readability scores from 3 formulas used. To measure readability, there are various formulas to choose from. The most common formula is the Flesch Reading Ease Score. The output of the Flesch Reading Ease is a number from 0 to 100, with a higher score indicating easier reading. This formula is calculated as follow:

READFRE = 206.835 - (1.015 x ASL) - (84.6 x ASW)

ASL is average sentence length, deduced from the number of words divided by the number of sentences. ASW is average number of syllables per word, or the number of syllables divided by the number of words. In Flesch Reading Ease, EO's score was 82.761, GT's was 82.861, and DeepL's was 83.891.

The next formula is Flesch-Kincaid Grade Level. Contrary to the previous formula, in Flesch-Kincaid Grade Level the higher the number, the harder it is to read the text. The grade levels range from 0 to 12. Here is the calculation:

READFKGL = (.39 x ASL) + (11.8 x ASW) -
15.59

ASL is average sentence length. It is deducted from the number of words divided by the number of sentences. ASW is average number of syllables per word; derived from the number of syllables divided by the number of words. In general, a text should generally have more than 200 words before the Flesch Reading Ease and Flesch-Kincaid Grade Level scores can successfully be applied. The 3 texts we attempted to investigate have more than 200 words each, so there was no doubt in applying this formula. EO scored 3.053, GT scored 3.098, and DeepL scored 2.992.

Coh-Metrix L2 Readability recorded EO's, GT's and DeepL's scores of 19.037, 21.085 and 21.432 respectively. No matter the formulas used, DeepL output showed the best performance.

GT's and DeepL's accuracy and readability

To address the first research question namely accuracy, we first turned to SAE J2450 result. As presented in Table 2 GT Total Score is 77 while DeepL is 29. In order to determine the final score and find out the error percentage relative to the text, this sum is then divided by the number of words in the text which is 491 (GT) and 485 (DeepL). The number of words in each text is calculated by using Word Count feature in Microsoft Word. GT's final score is 0.16. It signifies that statistically GT's translated output contains 16 % error. In other words, GT's output holds 84% correct value. Meanwhile, DeepL's final score is deducted from 29 (total sum of error) divided 485 (number of words in the text), resulting 0.06. It signifies that DeepL's translated output contains 0.06 % error. To say it differently, DeepL's output holds 99.04% correct value.

Afterwards, to determine whether these 2 MT outputs pass or fail in the quality assessment, we refer to CdT Rubric from European Translation Centre which sets percentage range of 0 -100% (Mateo, 2014: 80). A translation with 0 - 39 % score is unacceptable. One with 40 - 59 % is below standard. A translation having 60 - 79% is acceptable. A good translation should have 80 - 99% score.

If we looked back, both GT's and DeepL's correct values are above 80% correctness, exceeding minimum score for good translation criterion. So, on SAE J2450 it can be stated that the outputs from GT and DeepL passed the quality assessment grade with their score 84 and 99.04 respectively and that DeepL output was more accurate than GT. To confirm this result, we conducted statistical analysis based on the result of Coh-Metrix.

Statistical analysis

Since Coh-Metrix only evaluates indices of the linguistic features and discourse representations of a particular text only, a comparison must be made for GT, DeepL, and EO as the reference in order to find out how close GT and DeepL are to human text. A one-way ANOVA analysis was carried out with the help of SPSS 25. The hypothesis employed in this study was as follow:

- H0: $\mu1 = \mu2 = \mu3 = \ldots = \mu n$. There is no statistically different significance in mean of n groups.
- H1: $\mu1 \neq \mu2 \neq \mu3 \neq \ldots \neq \mu n$. There is a statistically different significance in mean of n groups.

To determine whether the differences between the means are statistically significant, the p-value was compared to the significance level to assess the null hypothesis. The null hypothesis states that the population means are all equal, which in this case meaning that all linguistic features of the text are equal. A significance level of 0.05 was used. P-value $\leq \alpha$ states that the null hypothesis is rejected and the differences between some of the means are statistically significant. If p-value $> \alpha$, then the differences between the means are not statistically significant.

Table 4: Descriptive Statistics

Score								
	N	**Mean**	**Std. Dev.**	**Std. Error**	**95% Confidence Interval for Mean**		**Min**	**Max**
					Lower Bound	**Upper Bound**		
EO	11	99.69	136.80	41.25	7.78	191.59	.47	463.0
GT	11	100.40	140.25	42.29	6.18	194.63	.46	472.0
DeepL	11	100.78	141.28	42.60	5.87	195.69	.45	479.0
Total	33	100.29	135.03	23.51	52.41	148.17	.45	479.0

As shown in the table that mean of these 3 corpora is not much different from one another. GT mean is 100.4037 and DeepL mean is 100.7793 and EO mean is 99.6848. GT's score is slightly lower than DeepL's but higher than EO. So, it suggests that overall GT output is closer to EO and better than DeepL. GT's and DeepL's standard deviations are bigger than EO. It indicates that these MT outputs are greater in score variation of their indices compared to EO. Or, GT and DeepL outputs are more variative in par. count, sent. count, word count, noun, verb, adjective and adverb incidences.

Table 5: ANOVA

Score					
	Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**
Between Groups	6.804	2	3.402	.000	1.000
Within Groups	583462.541	30	19448.751		
Total	583469.345	32			

ANOVA hypothesis can be described as follow:

H0 = there is no significant difference in mean of EO, GT and DeepL.

Ha = there is significant difference in mean of EO, GT and DeepL.

In order to draw a conclusion, we need F distribution value. With significance level 0.05 df Between Groups 2 and df Within groups 27, F Table is .051.

If F Statistic > F Table, then Ho is rejected.

If F Statistic < F Table, then Ho is accepted.

The table shows that F Statistic is .000, because .000 < .051, then Ho is accepted. It can be said that there is no statistically different significance between mean of EO, GT dan DeepL. Or, from probability value if p > 0.05, H0 is accepted. If p < 0.05, H0 is rejected. Since p value 1.000 is > 0.05, then H- is rejected meaning that there is no significant mean difference of EO, GT and DeepL.

Table 6: Post Hoc Tests

Multiple Comparisons						
LSD						
(I) Text_ Type	**(J) Text_ Type**	**Mean Difference (I-J)**	**Std. Error**	**Sig.**	**95% Confidence Interval**	
					Lower Bound	**Upper Bound**
EO	GT	-.7189	59.465	.990	-122.164	120.726
	DeepL	-1.0945	59.465	.985	-122.539	120.350
GT	EO	.7189	59.465	.990	-120.726	122.164
	DeepL	-.3756	59.465	.995	-121.820	121.069
DeepL	EO	1.0945	59.465	.985	-120.350	122.539
	GT	.3756	59.465	.995	-121.069	121.820

The P-value of GT (I Group) and EO (J Group) mean difference comparison is .7189, which is > than .05. So, there is not a statistically significant difference between GT output and human text EO. It goes the same way with DeepL to EO comparison. DeepL (I Group) and EO (J Group) mean difference comparison is 1.0945, which is also > than .05. So, there is not a statistically significant difference between DeepL output and human text EO. Since no

significant difference in mean score comparison between GT, DeepL and EO was found, then we can't decide which system is statistically better.

In terms of readability, DeepL proved to be the text offering the best reading ease as shown by the 3 readability formulas namely Flesch Reading Ease, Flesch-Kincaid Grade Level and Coh-Metrix L2 Readability which all favored DeepL.

CONCLUSION AND RECOMMENDATION

This study presented a quantitative assessment of Google Translate and DeepL translation outputs with French playscript *En attendant Godot* as the source text. Its objective was to find out how good was GT and DeepL in translating literary text from French to English. Manual assessment using SAE J2450 metric revealed that DeepL output is better than GT output in 7 primary error categories. DeepL scored higher than GT and consequently can be said to be more accurate.

Coh-Metrix assessment which analyzed text properties pertinent to descriptive aspects, word information, and readability discovered that these 3 corpora namely GT, DeepL and EO are overall similar yet having a variety. In terms of descriptive aspect both GT and DeepL are close to EO in 3 different indices respectively. In word information aspect DeepL is closer to EO compared to GT.

In terms of readability all 3 formulas are in favor of DeepL. It indicates that DeepL is easier to read and understand than GT. However, ANOVA Test demonstrates that these 3 corpora GT, DeepL and EO are not of significant difference. GT's mean of text properties score however is a little bit lower than that of DeepL but it doesn't imply that GT is better than DeepL. At first glance it can be said that GT is closer to EO because GT to EO mean difference is smaller than DeepL to EO. However, deeper investigation proved that this statistical difference only describes distinction in text properties and is probably due to DeepL different strategy of translating which resulted in bigger number of indices. Since these 3 texts are not significantly different from statistic point of view, so, it cannot be stated which system is better, GT or DeepL.

Despite this so-called statistics achievement, there are still many things that GT and DeepL need to improve. One noticeable example is lack of world knowledge which result in their failure to find proper equivalent in the target language especially for connotative and colloquial expressions; also, their inability to decipher ambiguity caused by structural differences between French and English. These are quantitatively not too meaningful but deep reading by human would find it disturbing.

For future research, it is strongly suggested to investigate the same topic for distant language pair like French to Indonesian or English to Indonesian. If Coh Metrix is used, then more indices and more types of corpora should be involved in order to ensure general validity and more comprehensive interpretation of the results. Regardless its limitations, this study has given rise to interesting findings yet debatable and needs further investigation: MT's output's proximity to human text, something that was previously deemed impossible.

REFERENCES

Ahrenberg, L. (2017). Comparing machine translation and human translation: A case study. *In Proceedings of The First Workshop on Human-Informed Translation and Interpreting Technology* (pp. 21–28). Wolverhampton, UK.

Arnold, E. A. 1994. *Machine Translation: An Introductory Guide*. London: Blackwells-NCC.

Bell, Roger T. (1991). *Translation and Translating: Theory and Practice*. London and New York: Longman.

Carroll., John B. 1966. *An experiment in evaluating the quality of translation. Mechanical Translation and Computational Linguistics*, 9(3-4):67–75.

Crystal, D. (1991). The Cambridge encyclopedia of language. Cambridge, UK: Cambridge University Press.

Finch, A., Hwang, Y.S. and Sumita, E. (2005). Using machine translation evaluation techniques to determine sentence-level semantic equivalence. In Proceedings of the Third International Workshop on Paraphrasing (IWP2005) (pp. 17-24) http://aclweb.org/anthology/I05-5003

Fowler, C. A., & Hodges, B. H. (2011). *Dynamics and languaging: toward an ecology of language. Ecological Psychology*, 23(3), 147-156. Retrieved on August 16, 2021 from http://dx.doi.org/10.1080/10407413.2011.591254

Graesser, A. C., McNamara, D. S., & Louwerse, M. M (2003). What do readers need to learn in order to process coherence relations in narrative and expository text. In A.P. Sweet and C.E. Snow (Eds.), Rethinking reading comprehension. New York: Guilford Publications.

Graesser, A. C., McNamara D. S., Louwerse, M. and Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. Behavior Research Methods, Instruments, & Computers 2004, 36 (2), 193-202.

Horguelin, Paul and Louise Brunette. 1998.Practique de la Révision, *3ème Edition Revue et Augmentée*. Québec: Linguatech éditeur.

Khomeijani, F. 2005. *A Framework for Translation Evaluation*. Cambridge: Blackwell Publishers Inc.

Larose, R. 1998. "*Méthodologie de l'Évaluation des Traductions".* [*A Method for Assessing Translation Quality*]. Meta 43: 163-86.

Larson, M. L. 1984. *A Guide to Cross language Equivalence*. Maryland: University Press of America.

Li, H., Graesser, A.C., and Cai, Z. (2014). Comparison of Google translation with human translation. In *Proceedings of the Twenty-Seventh International*

Martinez, R.M.2014. A Deeper look into metrics for translation quality assessment (TQA). *In Miscellanea: A Journal of English and American studies* 49 (2014): pp. 73-94 ISSN: 1137-6368

Nida, Eugene A. & Taber, Charles R. (1969). The theory and practice of translation. Leiden: E.J. Brill.

Nord, Christiane. 1997. *Translation as a Purposeful Activity*. Manchester, UK: St. Jerome.

O'Brien, Sharon. 2012. "*Towards a Dynamic Quality Evaluation Model for Translation*". Jostrans: The Journal of Specialized Translation 17. Retrieved from http://www. jostrans.org/issue17/art_obrien.php (Accessed 16 August, 2021) as cited in Dialnet

Parra Galiano, Silvia. 2005. La Revisión de Traducciones en la Traductología : Aproximación a la Práctica de la Revisión en el Ámbito Profesional Mediante el Estudio de Casos y Propuestas de Investigación. (*Doctoral Dissertation*, Universidad de Granada, Spain). Retrieved from http://digibug.ugr.es/handle/ 10481/660 (Accessed 31 July, 2021) as cited in Dialnet

Puchała-Ladzińska, K. (2016). *Machine translation: a threat or an opportunity for human translators? Studia Anglica Resoviensia* 13: 89-98.

Rahimi, R. 2004. *Alpha, Beta and Gamma Features in Translation: Towards the Objectivity of Testing Translation, Translation Studies*. Norwood: Ablex Publishing.

Richards, J., Platt, J., & Platt, H. (1997). Dictionary of language teaching and applied linguistics. London: Longman.

SAE. 2001. *Translation Quality Metric* [J2450]. Warrendale, PA: SAE.

Secâra, Alina. 2005. Translation Evaluation – a State of the Art Survey. *eCoLoRe/MeLLANGE Workshop Proceedings*. Leeds, UK: University of Leeds Press: 39-44.

Shankland, S. (2013). *Google Translate Now Serves 200 Million People Daily*. CNET. Retrieved from http://news.cnet.com/8301-1023_3-57585143-93/googletranslate-now-serves-200-million-people-daily/

Štefčík, J. (2015). *Evaluating Machine Translation Quality: A Case Study of Translation - a Verbatim Transcription from Slovak into German*. Vertimo Studijos. 2015. ISSN 2029-7033.

Sun, Sanjun. (2015). Measuring translation difficulty: Theoretical and methodological considerations. *Across Languages and Cultures*. 2015. DOI: 10.1556/084.2015.16.1.2. Retrieved from https://www.researchgate.net/publication/277922224 on 17 August 2021.

Templin M. C. (1957) Certain language skills in children. Minneapolis: University of Minne- sota Press

Torop, P. (2002). Translation as translating as culture. Sign System Studies, 30(2), 593–605.

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Lukasz Kaiser, Stephan Waddington, Christopher. 2000. *Estudio Comparativo de Diferentes Métodos de Evaluación de Traducción General* (Inglés–Español). Madrid, Spain: Universidad Pontificia de Comillas.

Waddington, Christopher. 2000. Estudio Comparativo de Diferentes Métodos de Evaluación de Traducción General (Inglés–Español). Madrid, Spain: Universidad Pontificia de Comillas.

Williams, Malcolm. 1989. “The Assessment of Professional Translation Quality: Creating Credibility out of Chaos”. TTR: Traduction, Terminologie, Redaction 2: 13-33.

APPENDIX

Tabulation Data of ST, EO, GT and DeepL

Source text (ST), in French, is *En attendant Godot*, a play written by Samuel Beckett. English Original (EO) was done by the author of the text himself in his English version *Waiting for Godot*. The translations of Google Translate (GT) and DeepL are both in English and were generated on July 31, 2021.

Remark:

1. Red represents Addition/Omission
2. Yellow represents Syntactic Error
3. Blue represents Wrong Term
4. Green represents Word Structure
5. Brown represents Punctuation

Source Text (in French)	English Original	Translation of Google Translate	Translation of DeepL
Route à la campagne, avec arbre. Soir. Estragon, assis sur une pierre, essaie d'enlever sa chaussure. Il s'y acharne des deux mains, en ahanant. Il s'arrête, à bout de forces, se repose en haletant, recommence. Même jeu. Entre Vladimir. **ESTRAGON** (renonçant à nouveau). - Rien à faire. **VLADIMIR** (s'approchant à petits pas raids, les jambes écartées). - Je commence à le croire. (Il s'immobilise.) - J'ai longtemps résisté à cette pensé, en me disant, Vladimir, sois raisonnable. Tu n'as pas encore tout essayé. Et je reprenais le combat. (Il se recueille, songeant au combat. A Estragon.) - Alors, te revoilà, toi. **ESTRAGON**. - Tu crois ? **VLADIMIR**. - Je suis content de te revoir. Je te croyais parti pour toujours. **ESTRAGON**. - Moi aussi. **VLADIMIR**. - Que faire pour fêter cette réunion ? (Il réfléchit.) Lève-toi que je t'embrasse. (Il tend la main à Estragon.) **ESTRAGON** (avec irritation). - Tout à l'heure, tout à l'heure. **Silence.** **VLADIMIR** (froissé, froidement). - Peut-on	.. (A) .. country road. A tree. Evening. Estragon, sitting on a low mound, is trying to take off his boot. He pulls at it with both hands, panting. He gives up, exhausted, rests, tries again. As before. Enter Vladimir. **ESTRAGON**. - (giving up again). Nothing to be done. **VLADIMIR**. - (advancing with short, stiff strides, legs wide apart). I'm beginning to come round to that opinion. All my life I've tried to put it from me, saying Vladimir, be reasonable, you haven't yet tried everything. And I resumed the struggle. (He broods, musing on the struggle. Turning to Estragon.) So, there you are again. **ESTRAGON**. - Am I? **VLADIMIR**. **-** I'm glad to see you back. I thought you were gone forever. **ESTRAGON**. - Me too. **VLADIMIR**. - Together again at last! We'll have to celebrate this. But how? (He reflects.) Get up till I embrace you. **ESTRAGON** (irritably). - Not now, not now. **Silence.** **VLADIMIR** (hurt, coldly). - May one inquire where His Highness spent the night? **ESTRAGON**. - In a ditch.	.(A)…Country road, with tree. Evening. Tarragon*, sitting on a stone, tries to take off his shoe. He persists at it with both hands, gasping. He stops, at the end of his strength, rests panting, starts again. Same game*. Enter Vladimir. TARRAGON (renouncing again). - Nothing to do. VLADIMIR (approaching with small, stiff steps, legs spread). - I'm starting to believe it. (He freezes.) I resisted that thought for a long time, telling myself, Vladimir, be reasonable. You haven't tried everything yet. And I resumed the fight. (He recollects himself, thinking of the fight. To Tarragon.) - So, there you are again. TARRAGON. - You think so? VLADIMIR. - I'm glad to see you again. I thought you were gone forever. TARRAGON. - Me too. VLADIMIR. - What to do to celebrate this meeting? (He thinks.) Get up and I kiss you. (He holds out his hand to Tarragon). TARRAGON (with irritation*). - All on time, on time*. Silence. VLADIMIR (crumpled, coldly). - Can we find out	..(A)..Road in the country, with tree. Evening. Estragon, sitting on a stone, tries to remove his shoe. He struggles with both hands, panting. He stops, out of strength, rests panting, and starts again. Same game*. Enter Vladimir. ESTRAGON (giving up again). - Nothing to do. VLADIMIR (approaching with small, stiff steps, legs spread). - I'm beginning to believe it. (He stands still.) I resisted this thought for a long time, telling myself, Vladimir, be reasonable. You haven't tried everything yet. And I would take up the fight again. (He collects himself, thinking about the fight. To Estragon.) - So, here you are again. ESTRAGON. - You think so? VLADIMIR. - I'm glad to see you again. I thought you were gone forever. ESTRAGON. – So did I. VLADIMIR. - What to do to celebrate this meeting? (He thinks.) Stand up so I can kiss you. (He holds out his hand to Estragon). ESTRAGON (with irritation*). - Just now, just now*. Silence. VLADIMIR (crumpled, coldly). - Can we know where sir* spent the night? ESTRAGON. - In a ditch.

savoir où monsieur a passé la nuit ? **ESTRAGON**. - Dans un fossé. **VLADIMIR** (épaté). - Un fossé ! Où ça ? **ESTRAGON** (sans geste). Par là. **VLADIMIR**. - Et on ne t'a pas battu ? **ESTRAGON**. - Si… Pas trop. **VLADIMIR**. - Toujours les mêmes ? **ESTRAGON**. - Les mêmes ? Je ne sais pas. **Silence.** **VLADIMIR**. - Quand j'y pense ... depuis le temps... je me demande... ce que tu serais devenu. . . sans moi... (Avec décision.) Tu ne serais plus qu'un petit tas d'ossements à l'heure qu'il est, pas d'erreur. **ESTRAGON** (piqué au vif). - Et après ? **VLADIMIR** (accablé). - C'est trop pour un seul homme. (Un temps. Avec vivacité.) D'un autre côté, à quoi bon se décourager à présent, voilà ce que je me dis. Il fallait y penser il y a une éternité, vers 1900. **ESTRAGON**. - Assez. Aide-moi à enlever cette saloperie. **VLADIMIR**. - La main dans la main on se serait jeté en bas de la tour Eifel, parmi les premiers. On portait beau alors. Maintenant il est trop tard. On ne nous laisserait même pas monter. (Estragon s'acharne sur sa chaussure.) Qu'est-ce que tu fais ? **ESTRAGON**. - Je me déchausse. Ça ne t'est jamais arrivé, à toi ? **VLADIMIR**. - Depuis le temps que je te dis qu'il faut les enlever tous les jours. Tu ferais mieux de m'écouter. **ESTRAGON** (faiblement). - Aide-moi ! **VLADIMIR**. - Tu as mal ? **ESTRAGON**. - Mal ! Il me demande si j'ai mal !	**VLADIMIR** (admiringly). - A ditch! Where? **ESTRAGON** (without gesture). - Over there. **VLADIMIR**. - And they didn't beat you? **ESTRAGON**. - Beat me? Certainly (.) they beat me. **VLADIMIR**. - The same lot as usual? **ESTRAGON**. - The same? I don't know. **Silence** **VLADIMIR**. - When I think of it . . . all these years . . . but for me . . . where would you be . . . (Decisively.) You'd be nothing more than a little heap of bones at the present minute, no doubt about it. **ESTRAGON**. - And what of it? **VLADIMIR**. - (gloomily). It's too much for one man. (Pause. Cheerfully.) On the other hand (,) what's the good of losing heart now, that's what I say. We should have thought of it a million years ago, in the nineties. **ESTRAGON**. - Ah stop blathering and help me off with this bloody thing. **VLADIMIR**. - Hand in hand from the top of the Eiffel Tower, among the first. We were respectable in those days. Now it's too late. They wouldn't even let us up. (Estragon tears at his boot.) What are you doing? **ESTRAGON**. - Taking off my boot. Did that never happen to you? **VLADIMIR**. - Boots must be taken off every day, I'm tired telling you that. Why don't you listen to me? **ESTRAGON** (feebly). - Help me! **VLADIMIR**. - It hurts? **ESTRAGON** (angrily). - Hurts! He wants to know if it hurts! **VLADIMIR** (angrily). - No one ever suffers but you. I don't count. I'd like to hear what you'd say if you had what I have. **ESTRAGON**. - It hurts?	where Monsieur* spent the night? TARRAGON. - In a ditch. VLADIMIR (amazed). - A ditch! Where? TARRAGON (without gesture). - Over there. VLADIMIR. - And we didn't beat you? TARRAGON. - Yes… Not too much. VLADIMIR. - Always the same? TARRAGON. - The same? I do not know. Silence VLADIMIR. - When I think about it ... since time* ... I wonder ... what would have become of you. . . without me ... (Decidedly) You'd be just a little pile of bones by now, no mistake. TARRAGON. (stung to the quick). - And after*? VLADIMIR (overwhelmed). - It's too much for one man. (Pause. Vividly.) On the other hand, what's the point of getting discouraged now, that's what I'm telling myself. You had to think about it ages ago, around 1900. TARRAGON. - Enough. Help me remove this crap. VLADIMIR. - Hand in hand we would have thrown ourselves down the Eifel tower, among the first. We wore handsome then*. Now, it's too late. We wouldn't even be allowed to go up there. (Estragon struggles on his shoe.) What are you doing? TARRAGON. - I take off my shoes. Has it ever happened to you? VLADIMIR. - Since the time I told you to take them off every day. You better listen to me. TARRAGON (weakly). - Help me! VLADIMIR. - Are you in pain? TARRAGON. – Bad*! He asks me if it hurts! VLADIMIR (with anger). - It is never but you who suffer! … (Moi) … I don't count. Yet I would like to see you in my place. Tell me about it.	VLADIMIR (amazed). - A ditch! Where? ESTRAGON (without gesturing). - Over there. VLADIMIR. - And they didn't beat you? ESTRAGON. - Yes... Not too much. VLADIMIR. - Still the same? ESTRAGON. - The same? I don't know. Silence. VLADIMIR. - When I think about it ... all this time ... I wonder ... what you would have become ... without me... (Decisively.) You'd be nothing but a little pile of bones by now, no mistake. ESTRAGON (Stung to the core). - And then what? VLADIMIR (overwhelmed). - It's too much for one man. (A moment. Briskly.) On the other hand, what's the use of getting discouraged now, that's what I tell myself. We should have thought of this a long time ago, around 1900. ESTRAGON. - Enough. Help me get this junk off. VLADIMIR. - Hand in hand we would have thrown ourselves down the Eifel Tower, among the first. We looked good then. Now it's too late. They wouldn't even let us go up. (Estragon struggles with his shoe.) What are you doing? ESTRAGON. - I'm taking off my shoes. Has that ever happened to you? VLADIMIR. - I've been telling you to take them off every day. You'd better listen to me. ESTRAGON (weakly). - Help me! VLADIMIR. - Are you in pain? ESTRAGON. - Pain! He's asking me if I'm in pain! VLADIMIR (angry). - You're the only one who ever suffers! I don't count. (Moi)I would like to see you in my place. You'd tell me about it. ESTRAGON. - Did it hurt? VLADIMIR. - Painful! He's asking me if it hurt!

VLADIMIR (avec emportement). - Il n'y a jamais que toi qui souffres ! Moi je ne compte pas. Je voudrais pourtant te voir à ma place. Tu m'en dirais des nouvelles. **ESTRAGON**. - Tu as eu mal ? **VLADIMIR**. - Mal ! Il me demande si j'ai eu mal ! **ESTRAGON** (pointant l'index). - Ce n'est pas une raison pour ne pas te boutonner. **VLADIMIR** (se penchant). - C'est vrai. (Il se boutonne.) Pas de laisser-aller dans les petites choses. **ESTRAGON**. - Qu'est-ce que tu veux que je te dise, tu attends toujours le dernier moment.	**VLADIMIR** (angrily). - Hurts! He wants to know if it hurts! **ESTRAGON** (pointing). - You might button it all the same. **VLADIMIR** (stooping). - True. (He buttons his fly.) Never neglect the little things of life. **ESTRAGON**. - What do you expect, you always wait till the last moment.	TARRAGON. - Did you have pain? VLADIMIR. - Bad*! He asks me if I was in pain! Tarragon (pointing the index finger). That's no reason not to button yourself. VLADIMIR (leaning in). - It's true. (He buttons himself up.) No carelessness in the little things. TARRAGON. What do you want me to tell you, you always wait until the last moment.	ESTRAGON (pointing to his index finger). - That's no reason not to button up. VLADIMIR (leaning in). - That's right. (Buttoning up.) No sloppiness in small things. ESTRAGON. - What do you want me to say, you always wait until the last moment.