Applied Linguistics Research Journal
ALR Journal 2020;4(5):84–104

Original Research
doi: 10.14744/alrj.2020.98700

Adequacy in Machine vs. Human Translation: A Comparative Study of English and Persian Languages

Mehrdad Vasheghani Farahani [1]

[1]Leipzig University, *Germany*

ABSTRACT

Adequacy as the extent to which translation corresponds to the source text has always been a salient feature of any machine translation service. For this reason, the main objective of this comparative study was to analyze and compare translation adequacy in machine vs. human translation from English into the Persian language. In other words, this study sought to investigate whether machine translation product was adequate enough as compared to human translation. For this purpose, seven various texts were selected and translated from Persian to English by two translation students as well as machine translation services including Google Translate and Bing Translation. In addition, a reference translation by two expert translators was prepared against which human and machine translations were compared and analyzed. Moreover, the Sketch Engine software was exploited to analyze the translations. The results of the study revealed that there was no statistically significant difference between human and machine translations as compared to the reference translation and that machine translations were adequately appropriate translations. The findings of this research might have useful implications for research in the field of machine translations. In addition, the results of this research have practical implications for those who are interested in doing research in machine translation.

Keywords: Translation adequacy; Machine translation; Human translation; Comparative study; Persian and English language pairs.

Corresponding Author: Mehrdad Vasheghani Farahani

Phone: +39-075-974998
e-mail: mehrdadfarahani1365@gmail.com

Article citation: Vasheghani Farahani, M. (2020). Adequacy in machine vs. human translation: A comparative study of English and Persian languages, *Applied Linguisics Research Journal, 4*(5): 84-104.

Received Date: February 9, 2020
Accepted Date: May 12, 2020
Online Date: September 5, 2020

Publisher: Kare Publishing

E-ISSN: 2651-2629

1. Introduction

Producing an adequate and acceptable translation which conveys the content materials of the source text into the target text has always been the preoccupation of the experts in translation (Gerber, 2012). This requires too much time for the translators. For this reason, Machine translations (MT) have found their ways into our lives. Machine translation is defined as a process in which message from one language is transferred to another language by means of software. Machine translation service has been improved within the past years, mostly because communication between nations has been drastically increased (Li, Graesser, & Cai, 2014). In other words, "modern machine translation services such as Google Translate and Microsoft's Bing Translator have made significant strides toward allowing users to read content in other languages" (Denkowski, 2015, p. 5). This propensity to exploit machine translation services is because they are easy to use and free of charge (Wilks, 2008).

Human translation as compared to MT is represented by some unique features. First, comparing to MTs, human translation is usually done in a very low-speed process. Furthermore, human translation is edited by human either the translator himself or the editor; however, MTs requires post-editing which is carried out by the human. In addition, in human translation, there is no interference of human being. Moreover, the common view is that machine translation is evaluated by comparing it to the human professional translation (Papineni, Roukos, Ward, & Zhu, 2002). In this regard, Delpech (2014, p. 42) states:

> MT evaluation has two purposes. First, it analyzes, during the development of an MT system, the impact of a system modification on the quality of the translations. Second, the evaluation enables us to compare the systems between them, usually during a broad evaluation campaign. Each of these purposes has a matching evaluation technique.

The inclination to use MT is basically due to fact that MTs are becoming more and more prevalent among end users and many are dependent on them for their translation needs (Li, Graesser & Cai, 2014). Indeed, this enhancement of MTs is because "translators who work in technical domains will be increasingly required to interact with MT (2009, p.1). Albeit these advances and services in the field of machine translation, its final production in the target language still needs to be desired as there are shortcomings and inconsistencies in it. This has caused the notion of machine translation quality assessment.

Assessing machine translation is regarded as a vital area of study for measuring the efficacy of the existing machine translations as well as improving the machine translations in future (Martindale & Carpuat, 2018). Among the different kinds of translation quality assessment, one specific kind is assessing translation adequacy in machine translation as compared to human translation (Fiederer & O'Brien, 2009). Translation adequacy is defined as "the measure of the quantity of information existent in the original text that the translated text contains; it indicates whether the output is a correct translation of the original sentence in the sense that the right meaning is communicated" (Ayegba & Osuagwu, 2015, p. 2). To put it in another way, translation adequacy refers to the extent to which the output conveys the same meaning or information as the input does. One important aspect of adequacy in MT is that how much of the source language translation has been preserved in the target language (Koehn & Monz, 2006).

Translation adequacy has always been subjected to various judgments, but there is one common agreement between the experts that translation adequacy is directed and is evaluated in comparison to source text (Chesterman, 2016). Munday (2012) has approached translation adequacy from norms perspectives; believing that translation is adequate when it is towards source language norms and restrictions. He, however, puts forward the idea that no translation is very adequate nor is it acceptable as they are in a continuum. Shuttleworth and Cowi (1997) believe that although adequacy is rather such a complex phenomenon to be defined and there is not a unanimous agreement on it, it is a concept which is more related to source language constraints. In addition, Carroll, Holman, Segura-Bartholomew, Bird, and Busby's (as cited in Tung, Kennedy & 2009) propose six steps for evaluating translation adequacy as "translation, assess clarity and equivalence, back translation, field testing, assess reliability, and interpretation" (p. 249).

In spite of all the advances, experts in this field have mostly mentioned that assessing the quality (adequacy) of machine translation services (Mirzaeian, 2007) requires more study; arguing that evaluating machine translation quality in terms of adequacy is regarded as an emerging field of inquiry in the academia which requires further investigation. In other words, assessing machine translation quality has always been an interesting and attractive area of inquiry which has received less attention by academia (O'Brien, 2012). Moreover, research in the field of translation adequacy needs more research in less researched language pairs like English- Persian (Moradi, 2015). With regard to the above-mentioned issues and due to the fact that machine translation adequacy assessment is a growing area of research which deserves more inquiry, this study was an attempt to assess, in a comparative design, translation adequacy in human translation vs. machine translation in an English-Persian context.

2.Literature Review

2.1.Theoretical Perspectives

Defined as a kind of translation "which is performed wholly or partly by computer "(Shuttleworth & Cowi, 2009, p. 99), the idea of developing machine translation tools has a long history. In fact, the idea of inventing machine translations dates back to the age of electronic computers and when the non-military computers in the 1940s began to emerge (Hutchins, 2005). From the very beginning, researchers in the field of machine translation put their efforts on scientific and technical texts translation as these genres required less literally and sophisticated textual structures such as cultural- bound references; however, after a while, researchers began to focus more on other types of texts as there was a need for them in the commercial context (Hutchins, 2010). In the 1950s to 1960s, researchers began to make use of trial and error approaches to MT. In the 1960s, USA put too much effort to develop MT to have access to the writings of Russian writings while at the same time, European countries emphasized on English French language pairs. In the 1970s and 1980s, the MT systems found their ways to commercial markets by designing new models such as TITUS in France and CULT in China. From the 190s to the present time, the rule-based models of MT as well as other approaches including statistical and Neuro-based ones have been dominated (Wilks, 2008).

Machine translations are usually of four types/approaches as statistical, rule-based, example-based and neural machine translation (Daniel & Martin, 2018). In the first model, there is a need to compile a parallel corpus as large as possible so that the program can statistically map the characters between two languages. In a rule-based approach to machine translation, there is a need to provide the computer with detailed knowledge about lexical, morphological, functional and syntactical information so that the program can distinguish these grammatical variations between the source and target languages (Hartley, 2009). In example-based machine translation, "sentence is translated by analogy. A number of existing translation pairs of source and target sentences are used as examples. When a new source sentence is to be translated, the examples are retrieved to find similar ones in the source" https://www.andovar.com/machine-translation/. In the neural based machine translation, "machine translation (NMT) is based on the paradigm of machine learning and is the newest approach to MT. NMT uses neural networks that consist of nodes conceptually modeled after the human brain." https://www.andovar.com/machine-translation/. Regardless of the type of approaches, "machine translation (MT) is intended to automate the core task, i.e. the production of a string of words that will count as a translation of the source text (ibid, p. 106-107).

For the speed, availability and cost savings, MT can be preferred by many end users. This has prompted the researchers to design various models and types for MT. However, quality assessment in MT is a very important issue for scholars in this field (Castilho, Doherty, Gaspari, & Moorkens, 2018).

2.2.Empirical Studies of the Machine Vs. Human Translation

In a comparative research, Shahahbi (2009) investigated the output quality of machine translation in an English and Persian context. For this reason, he compared two translation software namely as Pars Translator & Padideh Translator. He found out that generally Padideh translation was the best output in comparison to Pars Translator. Also, the major problems the Persian machine translation software had were in the domains of morphology, syntactic ambiguity and complex sentences.

Hebresha and Aziz (2013) embarked on new research on machine translation evaluation in Arabic classical texts and their English translations. Indeed, they were to see how one can improve machine translation in a rule-based approach. In this regard, they used Google translate and CMAT as machine translation tools and designed a set of rules according to the grammatical rules of the two languages. The results showed that CMAT had a far better quality than that of Google translate.

Li, et al. (2014) compared Google translation with human translation in terms of adequacy. For *this objective, they selected Chinese and English texts as their corpus of the study. They selected Chinese and Arabic Language pairs as they were native speakers of these two languages. The results showed that the Google translate as compared to human translation had the same level of formality and cohesion; showing that at the semantic level, Google translate had very similar to human translation. This research, however, showed that at the syntactic and grammatical level, Google translation still needs a lot to be desired.*

In order to study the translation efficacy from English into Persian, Afshin and Alaeddini (2016) embarked on a new research in which they analyzed verb tense in Google translation. Their corpus consisted of the Novel Oliver Twist written by Charles Dickens. *What they found was that despite some similarities, the overall translation of the Google MT was not adequate nor was it acceptable in translating verb tense from English into Persian.*

In another study, Abusa'aleek (2016) studied acceptability and adequacy of Islamic texts in MTs. For this purpose, five Islamic text-genres covered supplication, part of a sermon, pillars of Islam, Hadith (sayings of the prophet of Islam) from Arabic into English were selected as the corpus of the study. For analyzing the adequacy, 4 kinds of machine translations; that is to say, World lingo, Babylon translation, Google translate, and Bing translator were selected. The results of this study indicated that among the four software, Google translate is regarded as the best tool in terms of adequacy and acceptability with the readability of 20 %, fidelity of 18.35%, syntax, 19.17% and terminology of 20.82% as compared to the other systems.

In a more recent study, Ghasemi and Hashemian (2016) conducted a research to analyze on Google translate translation. They applied Keshavarz's (1999) model of error analysis to compare quality of machine translation output. They found out that active/passive voice errors were the most and least frequent errors.

In the same year (2016), Abu Sa'aleek ran a comparative study on the adequacy and acceptability of machine translation in translating Islamic texts. For doing so, he selected four different machine translations including World lingo, Babylon translation, Google translate, Bing translator. His corpus of the study included consist of five Islamic text-genres covered supplication, part of a sermon, pillars of Islam, Hadith (sayings of the prophet of Islam) from Arabic into English. The findings of this study could show that from among the four various MTs, Google translated functioned better in terms of adequacy and acceptability. The results also indicated that Google translated had better output in terms of (accuracy, suitability, and well-formedness) and sub-characteristics (syntax, terminology, reliability and fidelity).

Hakiminejad and Alaeddini (2016) did a comparative research with the aim of examining the efficacy of verb tense translation from English into Persian. For this purpose, four passages of the novel Oliver Twist were translated by Google Translate and then compared with that of a human translation. As the results showed the Google translate could not translate verb tense form English into Persian, failing to produce an adequate translation.

This short review literature indicated that assessing adequacy in MT is a research inquiry which deserves more research specially in an English –Persian language pairs which have been less studied despite the extensive use of MT.

3.Methodology

The design of this research was comparative in nature to compare adequacy in machine vs. human translation from English into Persian. Indeed, it was comparative as it set to compare the human translation with machine translation in terms of adequacy.

3.1.Reference Translation

Usually, in order to make a judgment on translation adequacy, there needs to have a reference translation which is thought to be flawless, adequate and acceptable. To this purpose, seven various texts written in Persian (Farsi) were selected. The genre of the texts was informative and included written mode in various fields including economics, management, medical sciences, law, environment preservation, basic sciences and engineering. The texts were selected from miscellaneous genres to ensure the matter of text diversification. In addition, the texts were selected from various genres to assure that they were replete with specialized terminologies and jargons. The texts were written by Persian native speakers who were expert in each field of inquiry. Two Persian –English translators who graduated in English Translation Studies at the university level and who had years of functioning as accredited and freelance translators (around 8 years) were requested to translate the texts from Persian into English to produce a reference translation. They were permitted to use dictionaries of any kind to ensure their translation quality.

3.2.Scale for Scoring Translation Adequacy of the Referenced Translations

For ensuring the translation quality of the referenced translation, the translations produced by expert translators were assessed. For assessing reference translation, the model proposed by Waddington (2001) was used. This model was used as its validity and reliability is already assessed. Waddington's model is divided into four scoring various rubrics. The one which was employed in this research was model C, as it was easier and was in line with the objective of this research. This model of translation assessment (Waddington, 2001) is holistic. In this model any translation is scored in 5 different levels and between 0 and 10; according to the guidelines provided to the translation evaluator. The range between 0-10 gives the freedom to the reviewer to give more scores to the translators with better results and lower scores to the translators with lower results. The results showed that the reference translations could meet the criterion proposed by the scale to a great extent.

Table 1. *Scale for Holistic Method C*

Level	**Accuracy of Transfer of ST Content**	**Quality of Expression in TL**	**Degree of Task Completion**	**Mark**
Level 5	Complete transfer of source text information, only minor revisions to reach professional standard	Almost all translation reads like a piece originally written in English. There may be minor lexical, grammatical or spelling errors.	Successful	9,10
Level 4	Almost complete transfer; there may be one or two insignificant inaccuracies; requires certain amount of revision to reach professional standard.	Large sections read like a piece originally written in English. There are a number of lexical, grammatical of spelling errors.	Almost Completely Successful	7,8
Level 3	Transfer of the general idea(s) but with a number of lapses in accuracy; needs considerable revision to reach professional standard	Certain parts read like a piece originally written in English, but others read like a translation. There are a considerable number of lexical, grammatical of spelling errors.	Adequate	5,6
Level 2	Transfer undermined by serious inaccuracies, through revision required to reach professional standard	Almost entire text read like a translation; there are continual lexical, grammatical or spelling errors.	Inadequate	3,4
Level 1	Totally inadequate transfer of ST content; the translation is not worth revising.	The candidate reveals a total lack of ability to express himself adequately in English.	Totally Inadequate	1,2

3.3.Machine Translations and Human Translations

As a comparative study, two translation software were selected to produce machine translation: Google translate software and Bing translate software. The reason for the selection of these two software was that they are the prevalent translation software used by translators and are freely accessible (Torkaman, 2013). In line with the machine translations, the texts were translated by two translation students (non-expert) as the human translation (three texts were translated by one translator and four texts by another one). Once the translations were done by human translators and machine translations, they were analyzed and compared separately to the reference translation to evaluate their adequacy.

3.4.Scale for Assessing Translation Adequacy

The need for a reliable, sufficient and clear method for assessing machine translation has caused the creation of various evaluation models (Denkowski, 2010). In line with the significant role that machine translation tools play in translation and due to their capacity in translating a large number of texts, evaluation methods for machine translation is dispensable (Aleek, 2016). Usually, assessing machine translation encompasses various aspects; namely as adequacy, fluency and fidelity and acceptability (Hovy, 1999). This research was, however, limited to translation adequacy and did not take the other two aspects into account. As far as translation adequacy in MTs is concerned, there are both qualitative and quantitative scales and criterion. However, the one which was used in this study was the set of criterion proposed by Specia, Hajlaoui, Hallett and Aziz. This model was preferred as it was conducive to conduct and easy to follow. Their model of translation adequacy is composed of the following criterion:

- the ratio of number of tokens in source and target and vice-versa
- the absolute difference between the number of tokens in source and target normalized by source length
- the ratio of percentages of numbers, content-/ non-content words in the source & target
- the absolute difference between the number of superficial constructions in the source andtarget: brackets, numbers, punctuation symbols
- the absolute difference between the number of PP/NP/VP/ADJP/ADVP/CONJP phrases

  in the source and target
- difference between the number of 'person'/' location'/organization' entities insource and target sentences.

In addition to the above-mentioned criteria, the following set of criterion (Delpech, 2014, p. 43) was used by machine- translation evaluators:

- the number of word n-grams that the translation to be evaluated and the reference

  translation have in common, for n between 1 and 4;
- the (word number) size differences between translation to be evaluated and

  reference translation

As far as this research was concerned, the texts were analyzed by applying both sets of criterion for having a better and more solid evaluation of machine translation adequacy.

3.5.Sketch Engine Software

Analyzing texts (even texts of small quantity) is an arduous and time-consuming job. For this reason, it is recommended that software is used for such a sophisticated task. In order to be able to look for specific information on the reference translation and the machine/ human translations and for the evaluation of them a corpus software was exploited as it could give the results in a very short period of time. From among various software, Sketch Engine was selected. Sketch Engine software is a windows- supported software used mainly in Corpus Linguistics. This program was designed by Lexical Computing Ltd. (https://www.sketchengine.co.uk/). It gives the researchers a wide range of opportunities such as specific word extracting, concordance lines, keywords in the context and collocation patterns (McGillivray & Kilgarrif, 2013). This software was exploited in this research owing to the fact that it was necessary to extract some information from the texts.

4.**Result**

For doing the practical aspect of assessing translation quality, first the reference translation as the target text was analyzed based on the criteria mentioned. Then the same procedures were applied in human translation and machine translations. It is worth noting that although the corpus of the study was created by various sub-corpora, in the translation assessment phase, they were all regarded as one unified corpus.

Correlation Coefficient

For ensuring the correlation of the two experts' translations, their scores were analyzed. Table 2 below represents the results.

Table 2. *Correlations between Translators*

		Rater1PosE	Rate2PostE
Rater1PosE	Pearson Correlation	1	.730*
	Sig. (2-tailed)		.016
	N	10	10
Rate2PostE	Pearson Correlation	.730*	1
	Sig. (2-tailed)	.016	
	N	10	10

*. Correlation is significant at the 0.05 level (2-tailed).

As the data in Table 2 shows, the significant 2-tailed between rater one and two was .018. Therefore, there was an acceptable index of correlation between two raters.

Table 3 to Table 6 represent the results obtained from assessing reference translation. The reference translation was carried out as a basis against which the machine translations and human translation were evaluated.

Table 3. *Basic Information of the Reference Translation*

Number of Sentences	Number of Words	Number of Tokens	Number of Tags
50	1252	1416	46

Table 3 demonstrates the basic information on the reference translation. As it can be seen, the reference translation contains 50 sentences, 1252 words, 1416 tokens and 46 tags; respectively.

As the theoretical framework assigns, for assessing translation adequacy, it is necessary to scrutinize and discriminate content words and non-content words. Basically, content words are those words and expressions that "which refer to a thing, quality, state, or action and which have meaning (lexical meaning) when the words are used alone" (Richard & Schmidt, 2010, p.126). They usually contain noun, verbs, adjectives and adverbs. On the other hand, non-content words (functional words) are the words "which have little meaning on their own, but which show grammatical relationships in and between sentences" (ibid). They usually contain conjunctions, prepositions, auxiliary verbs, articles and pronouns.

Table 4. *Content vs Non-content Words*

The ratio of percentages of numbers, content-/ non-content words in the reference translation	
Content Words	Non-content Words
657	549

Table 4 reveals the number of content and non-content words in the reference translation. As the data can show in table2, the number of content words which includes noun, verbs, adjectives and adverbs are 657; whereas the number of function words including conjunctions, prepositions, auxiliary verbs, articles and pronouns are 549.

Another category for assessing machine translation adequacy difference between the number of superficial constructions in the source and target: brackets, numbers, punctuation symbols.

Table 5. *Number of Superficial Constructions in Reference Translation*

Brackets	Numbers	Punctuation Symbols
24	50	120

Table 5 shows the absolute number of superficial constructions in the reference translation. As it is conspicuous, the reference translation contains 24 brackets, 50 numbers and 120 punctuation symbols; respectively.

The next criteria against which a translation is evaluated in the absolute difference between the numbers of PP/NP/VP/ADJP/ADVP/CONJP phrases found in the reference translation and the machine/ human translation. For this purpose, the various grammatical categories were analyzed in terms of quantity including verbs, nouns, adjectives, adverbs, prepositions, conjunctions, articles, pronouns and auxiliary verbs.

Table 6. *Absolute Difference between the Number of PP/NP/VP/ADJP/ADVP/CONJP Phrases in Reference Translation*

Verb	Noun	Adjective	Adverb	Proposition	Conjunction	article	Pronoun	Auxiliary
109	371	142	29	212	95	159	26	57

Table 6 indicates the absolute number of verbs, nouns and etcetera in the reference transition. As it is obvious from the data, the reference translation contains 109 verbs, 371 nouns, 142 adjectives, 29 adverbs, 212 prepositions, 95 conjunctions, 159 articles, 26 pronouns and 57 auxiliaries.

The difference between the number of persons. Locations and organizations is another category for evaluating adequacy in machine and human translation.

Table 7. *Difference between the Number of 'Person'/' Location'/Organization' Entities in Reference Translation*

Persons	Locations	Organizations
3	7	4

Table 7 represents the absolute number of persons, locations and organizations in the reference translation. As it is conspicuous, there are instances of person, 7 occurrences of locations and 4 instances of organizations in the reference translation.

As Delpech (2014) puts it, n-grams are important features for assessing machine translation adequacy. N-grams are defined by (McEnery & Hardie) as "a sequence of n elements (usually words) that occur directly one after another in a corpus, where n is to or more "(247). Table 8 shows the number of n-grams in the reference translation.

Table 8. *The Number of n-grams in Reference Translation*

Two Words	Three Words	Four Words and More
822	354	166

As it is seen from data in Table 8, the reference translation contained 63 two- word n-grams, 29 three-word n-grams and only four and more than four- word n-gram.

Human vs. Machine Translation Tools

After calculating and obtaining the results of the reference translation, the same procedures were taken separately on the human translation and the Google and Bing translation tools; respectively. Tables 9 to 12 represent the results.

Table 9. *Differences between four Translations in Terms of Basic Information*

	Reference		Google		Bing		Human	
	n	%	n	%	N	%	n	%
Number of Sentences	50	1.8	36	1.8	30	1.4	41	1.9
Number of Words	1252	45.3	847	43.5	953	45.1	949	43.7
Number of Tokens	1416	51.2	1022	52.5	1091	51.6	1138	52.3
Number of Tags	46	1.7	43	2.2	41	1.9	45	2.1

Table 9 represents the differences between the number of sentences, words, tokens and tags in four translations. As it is obvious, reference translation is composed of 50 sentences, 1252 words, 1416 tokens and 46 tags. In addition, Google translation contains 36 sentences, 847 words, 1022 tokens and 43 tags. Bing translation is composed of 30 sentences, 953 words, 1091 tokens and 41 tags. Human translation has 41 sentences, 949 words, 1138 tokens and 45 tags.

Table 10. *Basic Information of the Machine Translations vs. Human Translation*

	Number of Sentences	Number of Words	Number of Tokens	Number of Tags
Google	36	847	1022	43
Bing	30	953	1091	41
Human	41	949	1138	45

As the data show in table 10, the Google translation includes 36 sentences; while the Bing and Human translations contain 30 and 41 sentences; respectively. As far as the number of words is concerned, Google translate is composed of 847 words; whereas, the Bing and Human translations are composed of 953 and 949 words; respectively. When it comes to the number of tokens, however, Google translation is composed of 1022 tokens; as opposed to 1091 and 1138 tokens of Bing and Human translations; respectively. Regarding the number of tags, it is conspicuous that Google translation has 43 tags; whereas, Bing has 41 and Human translation has 45 tags.

Table 11. *The number of Content and Non-content Words in Four Translations*

	Reference		Google		Bing		Human	
	N	%	n	%	n	%	n	%
Content words	657	54.5	476	65.6	498	55.8	527	58.3
None-content words	549	45.5	250	34.4	395	44.2	377	41.7

Table 11 gives comparative information on the distribution and usage of content words (verbs, adjectives, nouns and adverbs) vs. non-content words (prepositions, conjunctions, articles, pronouns and auxiliaries'). As it is clear from the data, reference translation is composed of 657 content words and 549 non- content words. Google translation is composed of 476 content words; while Bing and Human translation are composed of 498 and 527 words; respectively. On the other hand, Google translate is comped of 250 non-content words, while Bing translation is composed of 395 and human translation contains 377 non-content words; respectively.

Table 12. *Number of Superficial Constructions*

	Brackets	Numbers	Punctuation Symbols
Google	22	41	75
Bing	22	42	74
Human	22	43	119

Table 12 suggests the comparative number of superficial contractions (brackets, numbers and punctuation symbols) in the three translations. As the data show, Google translate contains 22 brackets, 41 numbers and 75 instances of punctuation symbols. However, Bing translation software contains 22 brackets, 42 numbers and 74 punctuation symbols. When it comes to Human translation, it is clear from the data that, it contains 22 brackets, 43 numbers, and 119 punctuation symbols.

Table 13. *Absolute Difference between the Number of PP/NP/VP/ADJP/ADVP/CONJP Phrases*

	Verb	Noun	Adjective	Adverb	Proposition	Conjunction	Article	Pronoun	Auxiliary
Google	76	278	92	11	137	57	140	11	45
Bing	80	303	99	16	154	48	146	5	30
Human	80	309	123	17	172	92	111	14	37

Table 13 shows the distribution of various grammatical categories in the three translations. As the data represent, Google translate contains 76 verbs; whereas, Bing and Human translations have both 80 verbs. In the case of noun category, it is clear from the data that Google translate has 278 instances of noun while Bing and Human translations have 303 and 309 instances of nouns; respectively. For the adjectives, Google translate contains 92 instances, while Bing and Human translations contain 99 and 123 adjectives. Regarding the adverbs, Google translate has 11 adverbs, while Bing and Human translations have 16 and 17 adverbs, respectively. When it comes to prepositions, it is conspicuous that Google translation is composed of 137 prepositions, Bing and Human translations are composed of 154 and 172 ones, respectively. As far as the conjunctions are concerned, Google translation is composed of 57, Bing translation of 48 and Human translation of 92 conjunctions; respectively. In the case of articles, Google translate contains 140 articles; whereas, Bing and Human translation contain 146 and 111 ones. In the case of pronouns, Google translate is composed of 11 pronouns; while Bing and Human translations include 5 and 14 stances of pronouns. The last grammatical category is auxiliary which can be found 45 times in Google translation, 30 times in Bing and 73 times in Human translation.

Table 14. *Frequently Distribution of 'Person'/' Location'/Organization' Entities in Four Translations*

	Reference		Google		Bing		Human	
	n	%	n	%	n	%	n	%
Person	3	21.4	3	27.3	4	26.7	5	26.3
Location	7	50	6	54.5	7	46.6	8	42.1
Organization	4	28.6	2	18.2	4	26.7	6	31.6

Table 14 reveals facts on the distribution of person, location and organization entities in three translations. As it is understood, in terms of persons, Human translation has 5 entities followed by Bing and Google translation, respectively. For the Location, Human translation contains 8 entities; while Bing and Google translations have 7 and 6, respectively. For the organization, human translation contains 6 entities; whereas Bing and Google translations have 4 and 2, respectively.

Table 15. *Frequency Distribution of Basic Information in Four Translations*

	Reference		Google		Bing		Human	
	n	%	n	%	n	%	n	%
Number of Sentences	50	1.8	36	1.8	30	1.4	41	1.9
Number of Words	1252	45.3	847	43.5	953	45.1	949	43.7
Number of Tokens	1416	51.2	1022	52.5	1091	51.6	1138	52.3
Number of Tags	46	1.7	43	2.2	41	1.9	45	2.1

Table 15 shows the frequency distribution of basic elements in four translations. The Chi-square test results of the table 13 showed that there was no any statistically significant difference between three translations (Google translation p: 0.39, Bing translation: p: 0.64 and human translation p: 0.54) and the reference translation. In other words, all three translations were similar to the reference translation in terms of the number of sentences, words, tokens, and tags. However, based on the results of the P-value, it was revealed that the significant difference of the Bing translation, when compared to Google and human translation, was less. Also, as the data demonstrates, there was not any statistically significant difference between Google and Bing (P = 0.53), Google and human (P = 0.99) and Bing and human (P = 0.55).

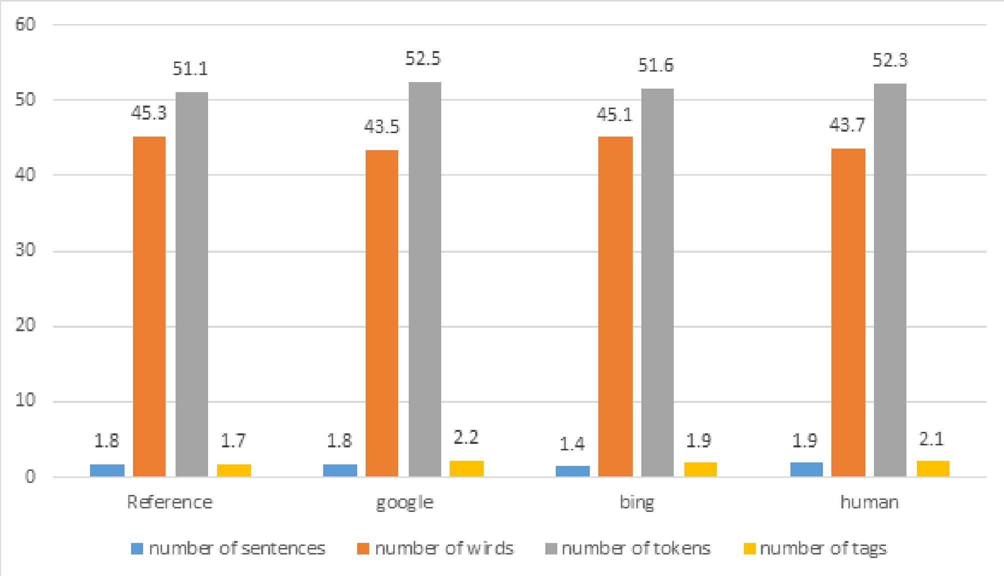

Figure 1. Frequently percent of basic information in four translations.

Figure 1 represents the distribution of sentences, words, tokens, and tags in four translations. As it is seen, in all translations, the highest number of distributive elements is dedicated to tokens with Google translate with the highest number (52.5) followed by the human (52.3), human (52.3) and reference translation (51.1). Regarding the number of words, reference translation contained the highest number of words with 45.3 followed by Bing (45.1), human (43. 7) and Google (43.5). Regarding the tags, as the data can show, as it can be seen, Google translate with 2.2 enjoyed the highest potation followed by Human (2.1), Bing (1.9) and reference with (1.7) as the least used category. For the number of sentences, Human translation contained the highest portion with 1.9 % followed by reference and Google translation both with 1.8. The least used was dedicated to Bing translation with only 1.4.

Table 16. *Frequently Distribution of Content/Non-Content Words in Four Translations*

	Reference		Google		Bing		Human	
	n	%	n	%	n	%	n	%
Content words	657	54.5	476	65.6	498	55.8	527	58.3
None-content words	549	45.5	250	34.4	395	44.2	377	41.7

Table 16 presents the distribution of content and non-content words in four translations. The results of the Chi-square test the frequency distribution of content words between Google translation and the Reference translation was significantly different (P < 0.001). In other words, the distribution of content words in reference translation was statistically significant. Also, the data showed that between Bing translation (P = 0.55) and Human translation (P = 0.08), there was no any statically significant difference. In other words, according to the calculated P-value, the difference of Bing translation and reference translation was less as compared to Google and Human translations. Also, there was a statistically significant difference between Bing translation (P < 0.001) and Human translation (P = 0.003); while between Bing and Human translation there was no any statistically significant difference (P = 0.28).

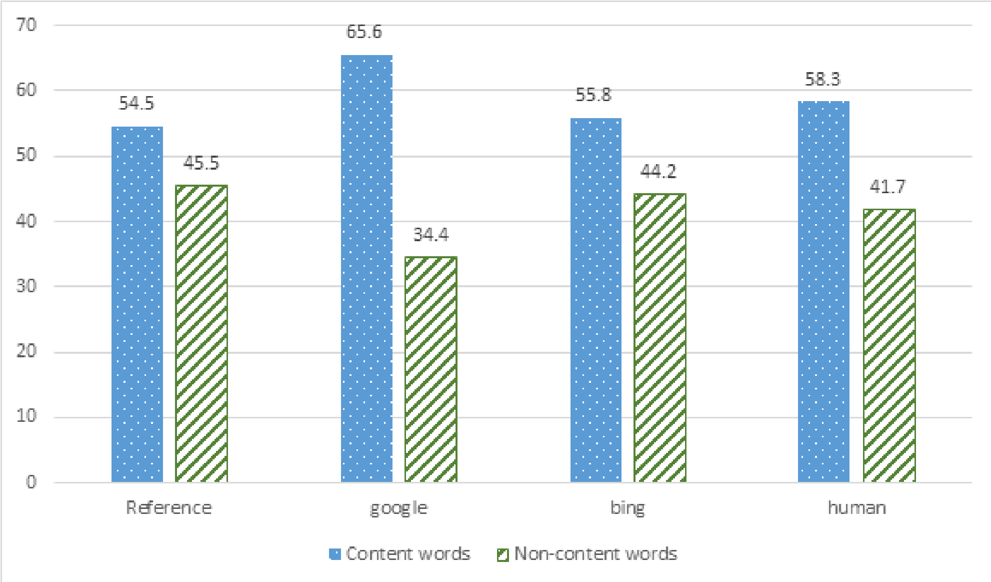

Figure 2. Frequency percentage of content/non-content words in four translations

As Figure 2 can show content words were used more than the non-content words in all translations. However, content words in Google translation (65.6) were more than the other translations followed by the human translation (58.3), Bing translation (55.8) and reference translation (54.5). Regarding the non-content words, reference translation enjoyed the highest percentage with 45.5 followed by the human translation (41.7), Bing (44.2) and Google translation (34.4).

Table 17. *Frequency Distribution of Superficial Constructions in Four Translations*

	Reference		Google		Bing		Human	
	n	%	n	%	n	%	n	%
Brackets	24	12.4	22	15.9	22	15.9	22	12
Numbers	50	25.7	41	29.7	42	30.4	43	23.4
Punctuation Symbols	120	61.9	75	54.4	74	53.6	119	64.6

Table 17 indicates the frequency distribution of superficial constructions (brackets, numbers and punctuation symbols). The results of the Chi-square test showed that there was no any statically significant difference between Google translation (P = 0.37), Bing (P = 0.31) and Human translation (P = 0.84) as compared to reference translation. In other words, all three translations were very close to reference translation. Nevertheless, based on the results, it can be understood from the P-value that from among the three translations, the Human translation was more close to the reference translation as compared to other translations. Furthermore, there was no any statistically significant difference between Google and Bing (P = 0.99), Google and Human (P = 0.17) and Bing and Human (P = 0.13).

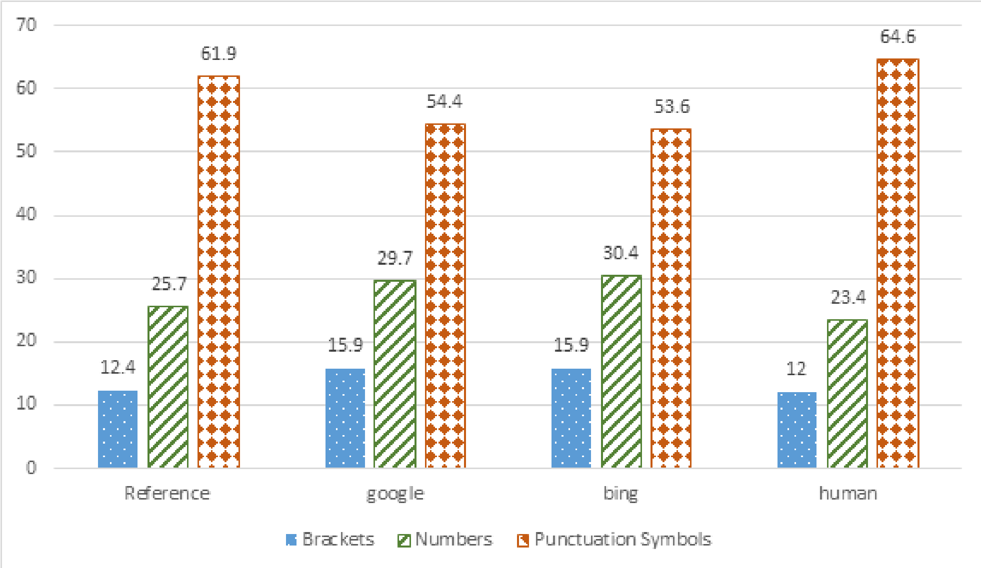

Figure 3. Frequency distribution of superficial construction in four translations.

Figure 3 represents the distribution of superficial constructions in four translations. As it is conspicuous, punctuation symbols were the most used superficial constructions in four translations with 61.9 % in reference translation, 64.6% in human translation, 54.4% in Google and 53.6% in Bing translations. After that numbers had the highest proportion of superficial constructions with Bing 30.4%, Google29.7%, reference 25.7% and human 23.4%. The least used superficial construction in four translations is bracketed with Google and Bing 15.9%, reference 12.4% and human 12%.

Table 18. *Frequency Distribution of PP/NP/VP/ADJP/ADVP/CONJP Phrases in Four Translations*

	Reference		Google		Bing		Human	
	n	%	n	%	n	%	n	%
Verb	109	9.1	76	9	80	9.1	80	8.4
Noun	371	30.9	278	32.8	303	34.4	309	32.4
Adjective	142	11.8	92	10.9	99	11.2	123	12.9
Adverb	29	2.4	11	1.3	16	1.8	17	1.8
Preposition	212	17.7	137	16.2	154	17.5	172	18
Conjunction	95	7.8	57	6.7	48	5.4	92	9.6
Article	159	13.2	140	16.5	146	16.6	111	11.6
Pronoun	26	2.2	11	1.3	5	0.6	14	1.5
Auxiliary	57	4.8	45	5.3	30	3.4	37	3.8

Table 18 demonstrates the frequency distribution of grammatical phrases in four translations. In terms of verbs, reference translation contained the highest number with 109 occurrences followed by Google (76), Bing (80) and human (80) occurrences. When it comes to the nouns, reference translation had the highest number with 371 occurrences. After that was the human translation with 309, Bing with 303 and Google with 278. For the adjectives, reference translation with 142 instances enjoyed the highest number followed by the human translation (123), Bing (99) and Google (92) instances. For the adverb, reference translation had the highest portion with 29 occurrences. After that human translation with 17 instances stood in the second place followed by Bing (16) and Google (11). In the case of the proposition, reference translation had the highest number with 212 occurrences. Then were the human translation with 172 items, Bing with 154 items and Google with 137 items. In terms of conjunctions, reference translation had the highest proportion with 95 times followed by the human translation (92), Google (57) and Bing translation (48). For the articles, reference translation with 159 times enjoyed the highest number of occurrences followed by Bing (146), Google (140) and human (111) instances. In the case of pronouns, the reference translation had the highest number of occurrences with 26 followed by human translation (14), Google (11) and Bing (5). For the auxiliary, reference translation contained 57 items as the highest proportion which followed by Google (45), human (37) and Bing (30).

The results of the Chi-square test showed that the frequency distribution of grammatical structures between reference translation and Google translation (P = 0.16) and human translation (P = 0.49) was not significantly different. However, for the Bing translation, there was a significant difference to reference translation (P = 0.005). Based on the results of the P-value, it can be said that the difference of human translation as compared to Google and Bing translation was less when compared to the reference translation. Also, there was not any statistically significant difference between Bing and Google translation (P = 0.36). However, there was a statistically significant difference between Google (P = 0.02) and Bing translations (P = 0.002).

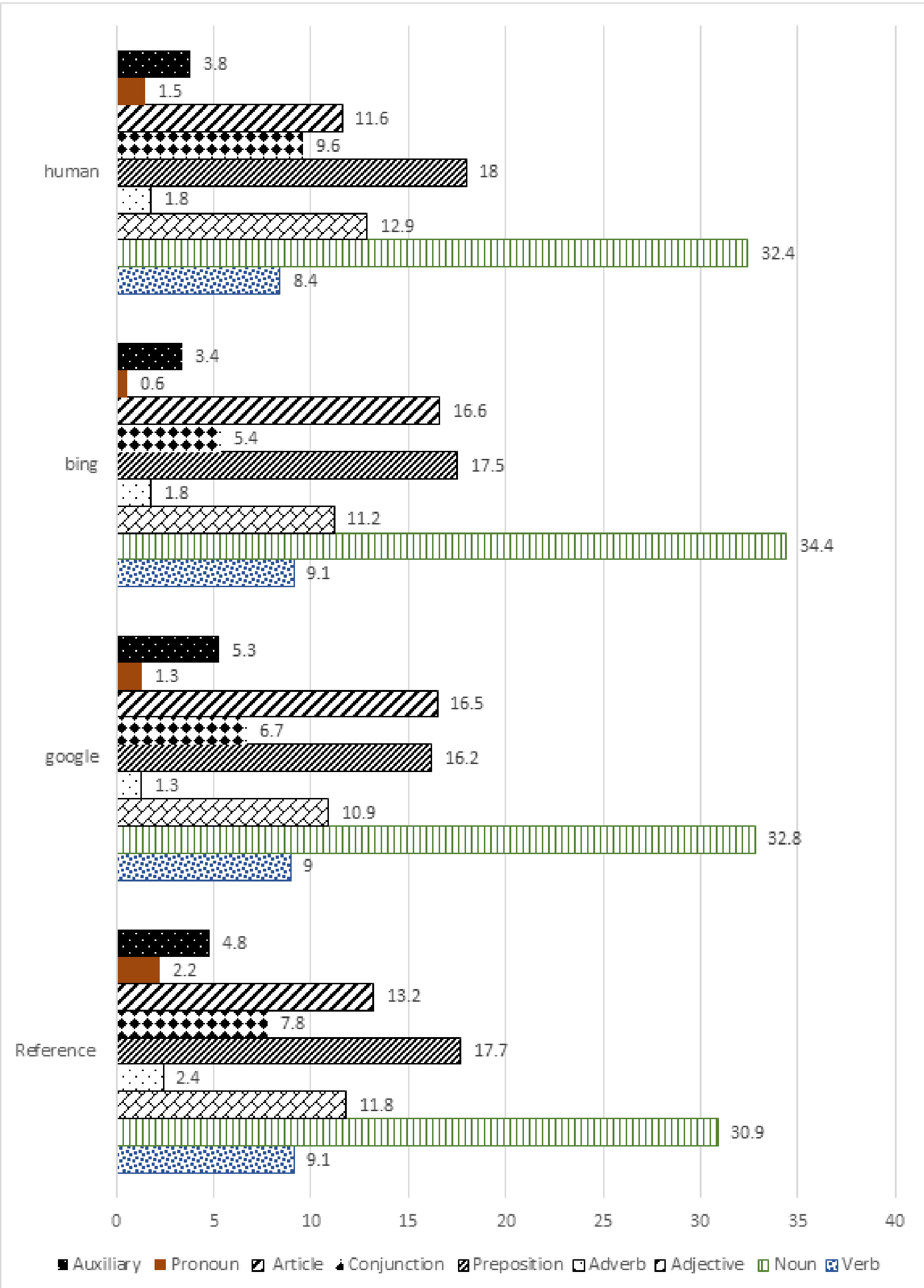

Figure 4. Frequency percent of PP/NP/VP/ADJP/ADVP/CONJP phrases in four translations

Figure 4 represents the percentage of the auxiliary, pronoun, article, conjunctions, prepositions, adverbs, adjective, nouns and verbs in four translations. As far as auxiliary is concerned, Google translation contained the highest number of auxiliaries (5.2%) followed by reference translation (4.8%), human translation (3.8%) and Bing translation (3.4%). Regarding the pronouns, reference translation contained 2.2% as the highest number followed by human translation (1.5%), Google translation (1.3%) and Bing (0.6%) as the lowest number of pronouns. For the articles, Bing and Google translations contained the highest number (16.6% and 16.5%, respectively). After that came reference translation and human translation with 13.2 % and 9.6%, respectively. Regarding the conjunctions, Google translation had the highest number with 5.3%, which was followed by reference translation (4.8%) and Google translation (3.8%). The lowest number of conjunctions was Bing translation with 3.4%. When it comes to preposition, it is clear from the data that human translation had the highest number of the proposition with 18%. After that was reference translation with 17.7%. Bing translation and Google translation were the next with 17.5 and 16.2%. In the case of the adverbs, reference translation contained the highest number of adverbs (2.4%). After that was human and Bing translations with both 1.8% percentage. The last one was Google translation with 1.3%. For adjectives, human translation contained the highest number (12.9%). After that, reference translation with 11.8% contained the highest number of adjectives, followed by Bing and Google with 11.2 and 10.9 %. Had the lowest number of adjectives. For the category of nouns, human and Bing translation with 32.4% contained the highest numbers which were followed by Google translation and reference translation with 32.8% and 309%, respectively. The last category was verbs. The reference and Bing translation contained the highest number of verbs with 9.1 %. After that were Google translation and human translation with 9% and 8.4%, respectively.

Table 19. *Frequency Distribution of 'Person'/' Location'/Organization' Entities in Four Translations*

	Reference		Google		Bing		Human	
	n	%	n	%	n	%	n	%
Person	3	21.4	3	27.3	4	26.7	5	26.3
Location	7	50	6	54.5	7	46.6	8	42.1
Organization	4	28.6	2	18.2	4	26.7	6	31.6

Table 19 depicts the frequency distribution of person, location and organization entities in four translations. The results of Chi-square test showed that the distributive frequency of a person, location and organization had no statistically significant difference with Google ($P = 0.82$), Bing ($P = 0.95$) and human translations ($P = 0.90$). In other words, all three translations were statistically close to the reference translation. However, with regard to P-value, it can be said that there was no any statistically significant difference between Google and Bing ($P = 0.87$), Google and human ($P = 0.70$) and Bing with human translations ($P = 0.95$).

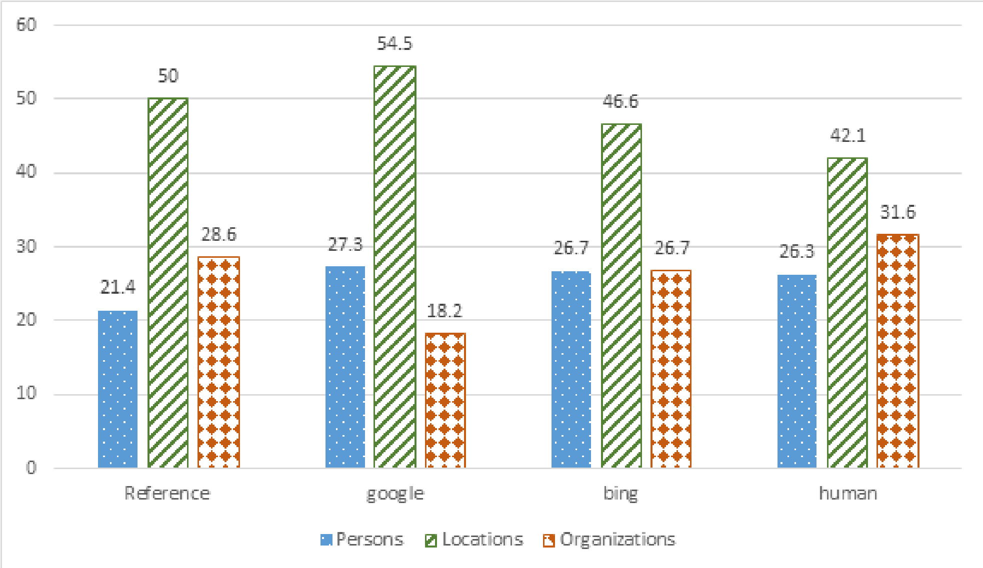

Figure 5. Frequency percentage of person, location and organization entities in four translations.

Figure 5 indicates the percentage of person, location and organization entities in four translations. As the data in table 6 represent, in all translations, location entity was the most prevalent one. However, in reference translation and human translation, the second place was dedicated to the organization entity, whereas Google translation, the person entity was the second most used entity. While, in reference and human translation, the person entity was the least used one, in Google translation, the organization was the least used entity.

Table 20. *Distribution of n-grams in Four Translations*

	Reference		Google		Bing		Human	
	n	%	n	%	n	%	n	%
Two words	822	61.2	493	52.9	648	55.1	1593	78.4
Three words	354	26.4	359	38.5	370	31.5	308	15.2
Four words or more	166	12.4	80	8.6	158	13.4	130	6.4

Table 20 demonstrates the distribution of N-grams in four translations. The result of the Chi-square showed that there was a statistically significant difference between Google ($P < 0.001$), Bing ($P = 0.006$) and human translations ($P < 0.001$). In other words, all three translations had a statistically significant difference as compared to the reference translation. Between Google and Bing, Google and human and Bing and human, there was a statistically significant difference ($P < 0.001$).

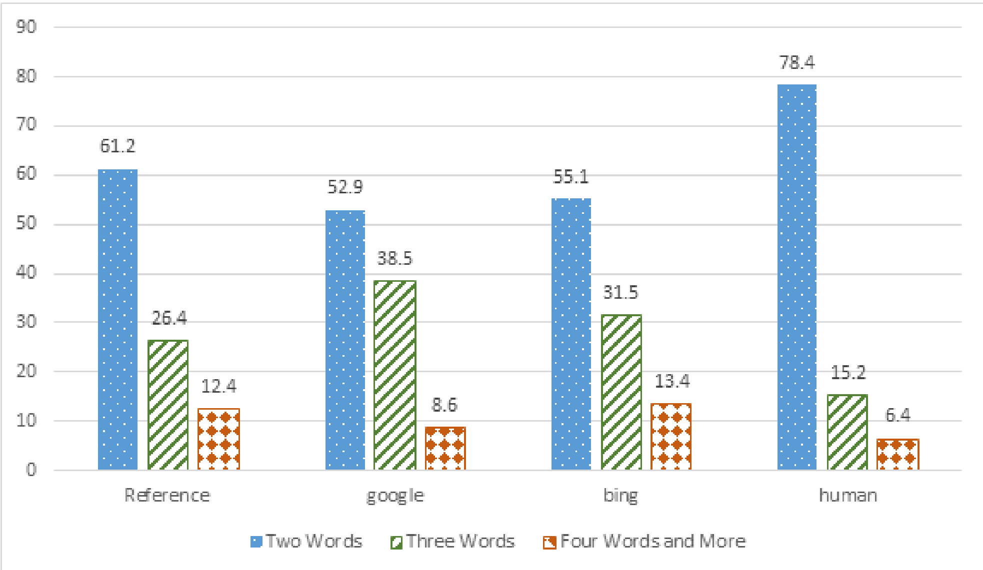

Figure 6. Distributive Percentage of n-grams in four translations

Figure 6 demonstrates the percentage of the n-gram in four translations. As it is clear, in all four translations, two-word n-grams were the most frequent one followed by three-word n-grams and four-word n-grams as the least used one in four translations.

Examples of the differences between the four translations

In order to clarify the differences between the three translations and the reference translation, it was necessary to bring authentic examples of translations from four resources.

Table 21. *Examples of Translations Performed by Machine Translation and Human Translation*

Reference Translation	**Bing Translation**	**Google Translation**	**Human Translation**
In order to maintain and restore aquifers, we need an accurate prediction of levels of underground water in different situations. (19 words).	In order to maintain and regenerate the vitality of the aquifer water level predictions need to be accurate in underground conditions. (21 words).	In order to maintain and rehabilitate the aquifer's life, it is necessary to accurately predict groundwater levels in different conditions. (20 words).	In order to maintain and revive the aquifer's life, it is necessary to accurately predict groundwater levels in different conditions. (20 words).
During the last decade, information technology has had a significant impact on the banking industry. This has enabled banks to provide their clients with distinctive products and services (27 words).	During the past decade, the role of dramatic effect Infor brsnat banking. This has caused banks to be able to mahswalt and متمایزی to service their clients (26 words, one Persian word).	Over the past decade, information technology has had a tremendous impact on the banking industry. This has enabled banks to provide distinct products and services to their customers (28 words).	Over the past decade, information technology has had a tremendous impact on the banking industry. This has enabled banks to provide distinct products and services to their customers (28 words).

5.Discussion and Conclusion

Online text translation services are becoming more and more prevalent due to their rapid functionality and versatility. Nowadays, most people are dependent on machine translation, as they have no knowledge of all languages. This research was conducted to compare the differences between machine translation and human translation in terms of adequacy. For this purpose, a corpus of seven texts written in various genres in Persian was selected. In addition, two expert English-Persian translators were requested to translate the texts into English to produce a reference translation. Once the reference translations were produced, two most popular machine translation services; that is to say, Google and Bing machine translations were employed to translate the same texts. In addition, the texts were translated by human translators to make a comparison between them and machine translations.

Regarding the first scale which is the number of content and non-content words in four translations, as the figure three represent, Google translation had the highest number of content words followed by Bing and Human Translation. On the other hand, with regards to non-content words, it was showed that Bing translation had the highest number of non-content words, followed by Human and Google translations.

Considering the second criterion, the distribution of superficial construction (brackets, numbers and punctuation symbols), as the figure number four can show the Google and Bing translations enjoyed an equal number (15.9) followed by the human translation which was 12. For the numbers, Bing translation contained the highest number (30.4) followed by Google and Human with 29.7 and 23.4%; respectively. However, in the case of punctuation symbols, Human translation with 64.6% functioned better than Google and Bing translations with 54.4 and 53.6%; respectively.

In terms of distribution of person, location and organization entities which is the fourth criteria for assessing translation adequacy in machine translation, as the data in table six demonstrates, Google translation enjoyed the highest number of persons (27.3%) followed by Bing (26.7%) and Human (26.3%). In the case of location, Google translation demonstrated the highest number of location entities (54.5%). After that were Bing and Human translation with 46.6% and 42.1%, respectively. When it comes to organization entity, Human translation with 31.6% depicted the highest number of entities followed by Bing and Google translate with 26.7% and 18.2%; respectively.

The last criteria for assessing translation adequacy was the number of N-grams. As the data in figure 7 can show, regarding the two- word n-grams, Human translation with 78.4% enjoyed the highest portion as compared to Bing and Google with 55% and 52%; respectively. For the three-word n-grams, as the data represent, Google translation with 38% ranked first, followed by Bing and Human with 31% and 15%; respectively. For the four- word and more n-grams, Bing translation with…13.4% enjoyed the highest number followed by Google and Human translations with 8.6 %and 6.4%; respectively.

Evaluating machine translation quality, like machine translation itself, has a long history (Hutchins 2001). This research was conducted to compare and contrast adequacy in machine translation vs. human translation. The quantitative results of the study could show that there was no statistically significant difference (s) between machine and human translations in terms of adequacy. In other words, machine translations could produce a translation which was very close to the reference translation. It can be said that as far as Persian- English language pairs are concerned, such translation software as Google and Being can produce an adequate output or translation. The final product, however, requires post-edit by an editor to compensate for inaccuracies of machine translation which can ensue. The results of this study are not in line with that of the *Li; Graesser & Cai (2014)* who proved that machine translation still needs to be desired to produce an acceptable translation into target language. Moreover, the results were in line with that of the Abusa'aleek (2016) study whose research showed that Google translate could produce an acceptable and adequate translation.

The results of this study can enlighten the differences between human translation and machine translation in terms of translation adequacy. This study is useful in that it can pave the way for the theoretical framework for machine translation adequacy. The results can also be used for software developers in the field of machine translations. They can use the results of this study for improving the adequacy in machine translations. The results can also give insights for doing comparative studies in the field of machine translation vs. human translation.

Machine translation research is not limited to adequacy; rather, other aspects of machine translation can be researched. Another research can be done in the field of machine translation fluency/naturalness. This research was focused on Persian (source text) – English text (target text). This research can be duplicated in an English _ Persian context as well as any other two language pairs. The current study was conducted based on a short piece of text. It is suggested that other studies with a longer stretch of texts are done to ensure the matter of generalizability.

References

Abusaaleek, A. O. (2016). The adequacy and acceptability of machine translation in

translating the Islamic texts. *International Journal of English Linguistics, 6(3), 185-193. Doi: 10.5539/ijel. v6n3p185.*

Afshin, H., & Alaeddini, M. A. (2016). A contrastive analysis of machine translation (Google Translate) and human translation: Efficacy in translating verb tense from English to Persian. *Mediterranean Journal of Social Sciences*, *7*(4), 40-48. doi:10.5901/mjss. 2016.v7n4s2p40.

https://doi.org/10.5901/mjss

Ayegba, S., & Osuagwu, O. (2015). Improving adequacy in a rule-based English-to-Igala Automatic translation system through word sense disambiguation. *British Journal of Mathematics & Computer Science*, *11*(1), 1-10. doi:10.9734/bjmcs/2015/19994.

https://doi.org/10.9734/BJMCS/2015/19994

Castilho, S., Doherty, S., Gaspari, F., & Moorkens, J. (2018). *Approaches to human and machine*

translation quality assessment (1st ed.). London: Springier.

Chesterman, A. (2016). *Memes of translation: The spread of ideas in translation theory*. Amsterdam: J. Benjamins.

https://doi.org/10.1075/btl.123

Daniel, J., & Martin, J. H. (2009). *Speech and language processing: An introduction to natural*

language processing, speech recognition, and computational linguistics. Englewood Cliffs, N: Prentice-Hall.

Delpech, E. M. (2014). *Comparable corpora and computer-assisted translation* (1st ed.).

London: Wiley.

https://doi.org/10.1002/9781119002659.ch1

Denkowski, M. (2010). *Choosing the right evaluation for machine translation: an examination*

of annotator and automatic metric performance on human judgment tasks. Paper presented at Proceedings of AMTA.

Denkowski, M. (2015). *Machine translation for human translators.* Unpublished doctoral

Dissertation. Pittsburgh.

Fiederer, Rebecca, & Sharon O'Brien (2009). Quality and machine translation: A realistic

objective? *The journal of specialized translation 11. 52-74. Web.*

Gerber, L. (2012). Machine Translation: Ingredients for productive and stable MT deployments –

Part 2. Retrieved from www.translationdirectory.com

Ghasemi, H., & Hashemian, M. (2016). A comparative study of google translate translations:

An error analysis of English-to-Persian and Persian-to-English translations. *English Language Teaching*, *9*(3), 13. doi:10.5539/elt. v9n3p13.

Hartely, T. (2009). Technology and translation. In Munday, J. (2009). *The Routledge companion to translation studies*. London: Routledge.

Hebresha, H. A., & AbAziz, M. J. (2013). Classical Arabic English machine translation using rule-based approach. *Journal of Applied Sciences*, *13*(1), 79.

https://doi.org/10.3923/jas.2013.79.86

Hovy, E. (1999). Toward finely differentiated evaluation metrics for machine translation. In

Proceedings of the EAGLES Workshop on Standards and Evaluation, 127-133. Pisa, Geneva. Munday, Jeremy. Introducing Translation Studies: Theories and Applications. London: Routledge Taylor & Francis Group, 2016. Print.

Hutchins, W.J. (2001). Machine translation over fifty years. *Histoire Épistémologie Language*, I (23):7–31. 2008. ILR language skill level descriptions. http://www.govtilr.org.

https://doi.org/10.3406/hel.2001.2815

Hutchins, J. (2005). Current commercial machine translation systems and computer-based

translation tools: system types and their uses. *International Journal of Translation*, *17*(1), 5-38.

Hutchins, W. J. (2010). Machine translation: A brief history. *Journal of Translation Studies, 13*,

29-70.

Koehn, P and C. Monz, C. (2006). Manual and automatic evaluation of machine translation between european languages. In Proc. Wkshp. *Statistical machine translation,* pages 102–121. https://doi.org/10.3115/1654650.1654666

Li, H., Graesser, A. C., & Cai, Z. (2014). *comparison of Google translation with human translation*. Paper presented at Twenty-Seventh International Florida Artificial Intelligence Research Society Conference, USA.

Martindale, M. J., & Carpuat, M. (2018, March). *Fluency over adequacy: A pilot study in measuring user trust in imperfect MT*. Paper presented at Proceedings of AMTA, Boston: USA.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory, and practice* (1st ed.). Cambridge: Cambridge Univ. Press. https://doi.org/10.1017/CBO9780511981395

McGillivray, B & Kilgrriff, A. (2013). Tools for historical corpus research and a corpus of Latin. In Bennett, P., Durrell, M., Scheible, S., & Whitt, R. J. (2013). *New methods in historical corpora*. Tübingen: Narr.

Moradi, M. (2015). *Comparison of two web-based translation machines in translating scientific-technical texts based on EuroMatrix model* (Unpublished master›s thesis). Kerman, Kerman.

O'Brien, S. (2012). Towards a dynamic quality evaluation model for translation. *The Journal of* specialized translation, (17), 55-77.

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: A method for automatic https://doi.org/10.3115/1073083.1073135 evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 311-318. Philadelphia.

Shahahbi, M. (2009). *An evaluation of output quality of machine translation program.* Paper submitted in Student research workshop, RANLP 2009 - Borovets, Bulgaria, pages 71–75.

Shuttleworth, M., & Cowie, M. (1997). *Dictionary of translation studies*. Routledge.

Torkaman, E. (2013). *A comparative study quality assessment of machine (Google translate) and human translation of proverbs from English to Persian* (Unpublished master›s thesis). Islamic Azad University: Central Tehran Branch, Tehran.

Waddington, C, (2001). Different methods of evaluating student translations: The question of validity, *Meta: Translators' Journal,* vol. 46, (2), 311-325.

Wilks, Y. (2008). *Machine translation: Its scope and limits* (1st ed.). Berlin, Germany: Springer Science & Business Media.