Machine Translation Quality Assessment of *Selected Works of Xiaoping Deng* Supported by Digital Humanistic Method

Qing Wang, Xiao Ma[*]

School of Foreign Languages and Literature, Wuhan University, Wuhan City, China

Email address:
839667471@qq.com (Qing Wang), 1486627668@qq.com (Xiao Ma)
[*]Corresponding author

To cite this article:
Qing Wang, Xiao Ma. Machine Translation Quality Assessment of *Selected Works of Xiaoping Deng* Supported by Digital Humanistic Method. *International Journal of Applied Linguistics and Translation*. Vol. 7, No. 2, 2021, pp. 59-68. doi: 10.11648/j.ijalt.20210702.15

Abstract: With the development of modern technology, the new technological methods need to be applied timely into translation research and translation quality assessment. As a new research method, digital technology is often used in the field of humanities research instead of the traditional ones and open a new paradigm for humanities research. At the same time, humanities research also provides a broad application space and rich corpus data for computer technology. Different from the traditional methods of machine translation quality assessment, this study attempts to apply the digital technology into the machine translation assessment, with the help of corpus technology, computer technology and statistical methods, so a to evaluate the quality of machine translations generated by different translation software from lexical, syntactical, semantic and pragmatic levels. The machine translation data for analysis come from the automatic translations by Baidu and Youda of the *Selected Works of Xiaoping Deng*, collection of Xiaoping Deng's important political speeches and theories, and the comparable data for reference is from the translated text produced by the expert translators. The specific case analysis and evaluation, on the one hand, verify the effectiveness of the digital humanity method applicable in the actual machine translation quality assessment; on the other hand, try to eliminate people's bias to the machine translation, so as to make people have a deeper understanding of the advantages and disadvantages of machine translation and improve the machine translation software design in the future.

Keywords: Digital Humanities, Machine Translation, Quality Assessment, Practicability

1. Introduction

In recent years, the machine translation system based on artificial neural network has made initial progress in simulating the neural network model of human brain to realize translation activities, but the computer still can not have the ability of thinking and reasoning like human beings, and the operational mode of computer translation is not the same as that of human translation. Therefore, the results of machine translation can not be completely evaluated by human translation standards. Moreover, the simple theory guided manual scoring evaluation method is too heavy and difficult to meet the needs of judging the results of machine translation in a short time. Inspired by "Digital Humanities", a new research paradigm dedicated to the application of modern computer technology to the research of traditional humanities [1], the author introduces digital technology into

the quality evaluation method of machine translation, so as to realize the integration of computer, linguistics and translation theories, and improve the objectivity and accuracy of the quality evaluation results of machine translation. Thus, the efficiency of machine translation quality evaluation has also been improved. Taking the Chinese English Parallel Corpus of *Selected Works of Xiaoping Deng* as the evaluation reference, this paper attempts to reflect the actual translation situation of the machine translation system based on the results of text data processing and analysis at the lexical level, sentence level, semantic level and pragmatic level with the help of human digital tools, and further verifies the accuracy and feasibility of the evaluation model.

2. The Existence of Translation Quality Evaluation Methods

Translation standards play a role in measuring the quality of translation in translation practice [2], which is a rule of conduct that must be followed in different translation activities, including manual translation and machine translation. The study of translation quality evaluation began in the 1970s, and most of them focused on the construction of the evaluation criteria of human translation quality. The existing evaluation models only stay in the theoretical stage and lack of applicability in machine translation assessment. However, the evaluation of machine translation quality emphasizes more about the pertinence, practicability and validity of the selected parameters within the scope of effective evaluation, and mainly focuses on the two aspects of faithfulness and fluency, which correspond to Lu Xun's "faithfulness" and "smoothness" standards [3].

2.1. Traditional Evaluation Model of Translation

Any translation needs quality assessment. Translation critics, reviewers, teachers, translation enterprises, experts and readers will judge translation in one way or another. At present, there are three major orientations in translation quality assessment: 1) source text oriented evaluation model. It is mainly to compare the translation with the original to see whether the translation is consistent with the original in some or all aspects. That is to say, the quality of the translation is mainly measured by the degree of similarity between the translation and the original. In other words, the higher degree of similarity between the translation and the original, i.e. the translated text is closer to the original, the better the quality of the translation [4]; 2) target text oriented evaluation model. It pay attention to whether the translation conforms to the cultural norms of the target language. In this vein, people usually compare the translated text with the similar text in the target language, namely, parallel texts, to see whether the translated text conforms to the target similar text norms. Its characteristic is that it no longer puts the original text in the supreme position, but pays more attention to the acceptability of the translation. And more considerations are given to the target readers, the norms of the target language and the social and cultural constraints. The evaluation criteria focus more on the degree of similarity between translation and similar texts in the target culture [5]; 3) the effect oriented evaluation model. Questionnaire survey, interview, comprehension test and translation comments are often used to evaluate the effect of translation and the criterion is the principle of equivalent effect [6].

In terms of quality assessment of machine translation, evaluation tends to be descriptive from the perspective of the nature of translation, and the evaluation parameters and their weights vary with different interpretations of translation [7]. However, in the academia, even the concept of translation is often defined differently in the field of translation, due to different understandings and choices of the connotation of the concept [8]. Roughly, the understandings of translation can be summarized into three categories: first, translation is regarded as a tool, that is, translation is taken as a means or a tool to achieve the purpose of communication and expression. Jacobson thinks that the same way of meaning expression in one language can be found in another language, that is, the ability of language expression is the same; second, translation is considered as creation. For example, Ezra Pound advocates "interpretative translation" and the freedom and creativity in poetry translation [9], and Yu Guangzhong regards translation activities as "limited creation" by "using Chinese to trap English who refuses to comply" [10]; third, translation is conceived as a kind of social activities, functioning in a specific social context, as in Hatim and Mason's opinion that translation is a communicative process which takes place within a social context.

2.2. Evaluation Model of Machine Translation

The transformation of machine translation from rule-based, corpus based to artificial neural network-based is to achieve automatic and high-quality artificial intelligence translation. However, at present, the quality of machine translation can not be compared with that of professional translation. Computer system only simulates the operation mechanism of human brain to process and produce information, which is far from the ability of human brain to judge and reason according to contextual information. In view of the limitations and special operation model of machine translation system, the evaluation parameters of machine translation are limited to the basic translation criteria of "faithfulness" and "smoothness".

2.2.1. Manual Evaluation Method

It is flexible but time-consuming to evaluate the quality of machine translation according to the standards and parameters set manually. The manual evaluation method generally takes the original text as the reference, and usually uses fidelity, fluency, intelligibility and information as the indicators. Each indicator is divided into several evaluation levels according to the situation of the translation. In practice, these four indicators are usually distinguished. The first two indicators are a group of evaluation parameters, and the second two indicators are another group of evaluation parameters. Faithfulness is the basic criterion of translation evaluation, and the goal of faithfulness is to reproduce the content, thought, emotion and style of the original text. But absolute faithfulness can not exist, and relative faithfulness is easier to be accepted and implemented. Fluency is used to evaluate the degree to which a translation adapts to the expression habits of the target language. The difference between intelligibility and fluency lies in the acquisition, processing and production of semantic knowledge blocks. According to the degree of understanding, it can be divided into four levels: complete understanding, basic understanding, less understanding and complete incomprehension. Information degree is a parameter to evaluate the degree of information the reader obtains from the translation. The larger the weight is, the richer the amount of information is, and the better the effect of machine

translation is. At present, researches in mainland China on the quality evaluation methods of machine translation include questionnaire and data analysis (see [11]), the quality estimation algorithm of machine translation based on translation knowledge (see [12]), the quality evaluation method of machine translation based on pretraining language model (see [13]), methods of quality evaluation of machine translation based on knowledge transfer (see [14]), AHP fuzzy evaluation method of machine translation quality in English translation (see [15]), translation quality estimation method based on subword sentence level neural machine translation (see [16]), translation quality estimation method based on word vector of BERT context (see [17]), the translation quality estimation method based on multilanguage pretraining language model (see [18]), etc.

2.2.2. Automatic Evaluation Method

According to the specific rules and algorithms, the automatic evaluation method can automatically score and sort the machine translation quality on the computer system. At present, the principle of machine translation automatic evaluation system is based on N-element matching, editing distance and error rate. Representative evaluation algorithms are BLEU, NIST, WER, PER, TER. BLEU and NIST calculate the similarity between machine translation and reference translation based on N-element grammar. The higher the similarity, the better the fidelity of machine translation. These two automatic evaluation methods help to enhance the objectivity of manual evaluation, and have certain reference value in machine translation evaluation [3]. WER is to calculate the number of times of editing (including insertion, deletion, movement and replacement) when the machine translation is converted into a reference translation. The less the number of times, the better the output of the machine translation system. PER and TER use another common evaluation method of machine translation, which is to identify the error rate in machine translation, formalize the specific translation errors and types, and judge the quality of machine translation from the error degree of analysis data. The method of machine translation quality evaluation based on similarity and error rate is only evaluated from the grammatical level. In order to investigate the quality of machine translation in a deeper level, the method based on linguistic knowledge (morphology, syntax, semantics, pragmatics, etc.) is introduced into the automatic evaluation system of machine translation. For example, Professor Shiwen Yu from the Institute of Computational Linguistics of Peking University proposed an automatic evaluation method of machine translation based on test points (such as word collocation, word order, syntactic structure, etc.). In recent years, the application of context in the computer field has aroused widespread concern. Linguists and machine translation system developers try to introduce the relevant theories of pragmatics into machine translation and machine translation evaluation.

Another important factor in automatic evaluation is the setting of evaluation parameters and weights. Evaluation parameter is the value or element that can be used for reference when evaluating something, and it is a variable. Setting weights for evaluation parameters is to determine the importance of each parameter in the evaluation system, which is expressed by weight coefficient. The larger the coefficient is, the larger the proportion of the parameter in the evaluation results is. Machine translation realizes the transformation between two natural languages based on language model and translation model. Therefore, language parameters (such as vocabulary, syntax, word order, semantics, pragmatics, etc.) are indispensable and occupy a large proportion in the automatic evaluation system of machine translation. Secondly, nonverbal parameters (such as culture, emotion, thinking, common sense, etc.) play an increasingly important role in machine translation quality evaluation. In general, the sum of the weights of each parameter is set to 1, and the weight of each parameter is expressed as a number from greater than 0 to less than 1 according to the degree of importance. The common methods of weight setting include analytic hierarchy process, fuzzy method, expert experience method and so on. Among them, analytic hierarchy process divides multiple levels of indicators, compares the importance of indicators by constructing judgment matrix, and compares the importance of indicators according to 9 importance levels. According to the theory of fuzzy mathematics, fuzzy method is a method to judge the weight of parameters; expert experience method is a method to synthesize the weight of each index according to the experience and subjective judgment of most experts.

At present, the statistical methods based on the co-occurrence of n-gram are BLEU (Bilingual Evaluation Understudy) and NIST (National Institute of Standards and Technology). The difference between them is that the former calculates the score according to the geometric mean of n-gram accuracy, while the latter calculates the score according to the arithmetic mean of n-gram accuracy. The higher the score, the better. Their formulas and algorithms are as follows,

BLEU:

$$BP = \begin{cases} 1 \ if \ l_c > l_s \\ e^{1-\frac{l_s}{l_c}} \ if \ l_c \le l_s \end{cases}$$

$$P_n = \frac{\sum_i \sum_k \min\left(count_k(c_i), max_{j \in m} count_k(s_{ij})\right)}{\sum_i \sum_k \min\left(count_k(c_i)\right)}$$

$$BLEU = BP \times \exp\left(\sum_{n=1}^{N} w_n log P_n\right)$$

Among them, l_c indicates the number of words in the candidate MT translation, l_s indicates the number of words in the standard reference translation, BP denotes penalty factor. When the number of words in the candidate MT translation is more than that in the standard reference translation, supposing the value of the penalty factor BP is 1, there is no penalty. When the number of words in the MT candidate translation is less than or equal to the number of words in the standard reference translation, punishment is required. $count_k(c_i)$ indicates the number of expression

words k occurring in the candidate machine translation, and $count_k(s_{ij})$ indicates the number of expression words k occurring in the standard reference translation. N is the value of n-gram, w_n represents weight, and the method of uniform weighting is generally adopted, that is, $w_n = 1/N$.

NIST:

$$Info(W_n) = log_2 \left(\frac{count(W_{n-1})}{count(W_n)} \right)$$

$$NIST = \sum_{n=1}^{N} \left\{ \sum_{W_n} Info(W_n) / \sum_{W_n} (1) \right\}$$

$$\cdot \exp \left\{ \beta log^2 \left[min \left(\frac{l_c}{l_s}, 1 \right) \right] \right\}$$

Among them, $count(W_{n-1})$ indicates the number of times n-1-gram appearing in the standard reference translation, and $count(W_n)$ indicates the number of times n-gram appearing in the standard reference translation. When $n = 1$, the value of $count(W_{n-1})$ is the number of all words in the standard reference translation. $Info(W_n)$ represents the amount of information of per n-gram. l_c indicates the number of words in the candidate MT translation, and l_s indicates the number of words in the standard reference translation.

The automatic evaluation method based on editing distance is mainly to calculate the number of editing (including insert, delete, move and replace) when machine translation is converted to reference translation. The less the number, the closer the level of machine translation is to that of manual translation. The minimum editing distance is a similarity calculation function. The smaller the editing distance is, the greater the similarity between the system translation and the reference translation is. WER (word error rate) is based on Levenshtein Distance. In the evaluation of machine translation results, we use words instead of strings as the calculation unit. The basic calculation formula is as follows,

WER:

$$WER = 100 \cdot \frac{S + D + I}{N} \%$$

Among them, S indicates the number of times the word was replaced, D indicates the number of times the word was deleted, I indicates the number of times the word was inserted, and N means the number of words. The calculation principles of PER, TER and WER are similar, that is, calculating the editing distance between the candidate machine translation and the annotated reference translation, but the difference is that PER ignores the word order in machine translation, that is, it allows the change of word position, while WER and TER consider the word position order in machine translation, and TER also supports block movement.

The automatic evaluation method based on linguistic knowledge investigates the accuracy and recalling rate of machine translation from stemming and synonym matching. The representative evaluation method is METEOR.

METEOR:

$$Pen = \gamma \left(\frac{ch}{m} \right)^{\theta}$$

$$F_{mean} = \frac{P_m R_m}{\alpha P_m + (1 - \alpha) R_m}$$

$$P_m = \frac{|m|}{\sum_k count_k(c_i)}$$

$$R_m = \frac{|m|}{\sum_k count_k(s_{ij})}$$

$$METEOR = (1 - Pen) F_{mean}$$

Among them, α, γ and θ represent the parameters of the model, m represents the word set given in advance based on WordNet thesaurus, ch represents a chunk formed by aligned and continuous words between the candidate machine translation and the standard reference translation, $count_k(c_i)$ indicates the number of times the expression word occurring in the candidate machine translation, and $count_k(s_{ij})$ indicates the number of times the expression word occurring in the standard reference translation.

With the emergence of large-scale corpus and the emergence of various machine translation algorithms, automatic evaluation has become an indispensable evaluation method in the development process [19]. To a great extent, the automatic evaluation model saves the time and energy of manual evaluation. It evaluates the machine translation through the concepts of similarity, error rate, accuracy and recalling rate, and realizes the automation, algorithmization and accuracy of the quality evaluation of machine translation, but it also has certain limitations. The disadvantage lies in that the evaluation strategies and methods can not be adjusted according to the characteristics and results of machine translation.

3. Case study of Machine Translation Evaluation Based on *Selected Works of Xiaoping Deng*

The quality assessment model of Chinese-English machine translation based on digital humanities method aims to evaluate the quality of machine translation from the linguistic level and non linguistic level with the help of corpus, python programming language, similarity algorithm and text sentiment analysis technology, etc. At the linguistic level, the consistency between the machine translation and the reference translation is evaluated from the parameters of functional words, keywords, word collocation, average sentence length and word order; at the non linguistic level, the semantic and contextual effects of the machine translation are judged by text information extraction and calculating the similarity and emotional inclination between the machine translation and the reference translation.

3.1. Evaluation at the Lexical Level

Vocabulary is the basic unit of machine translation

evaluation, which can be a single word or a group of words. In view of the grammatical features of terms and the semantic relationship between terms, the lexical level of MT evaluation starts from functional words and collocations. The number of functional words in Chinese is small, but the frequency of use is very high, and they must be attached to the notional words to have practical meaning. Whether they are used correctly or not directly affects the meaning of the whole sentence. Whether the machine translation system can recognize the functional words in the original text and translate them reasonably according to the context is related to the fidelity of the machine translation. Chinese and English belong to two different language families with different grammatical rules. Collocation is one of the most common grammatical rules. Firth first mentioned the concept of word collocation in *Modes of Meaning* [20]. He believed that the semantic meaning of a word is determined by another word that is collocated with it. In addition to the semantic association of collocation, other related studies have mentioned that there is a certain dependency between word and word combination, which may be due to the conceptual elements such as idiomatic language, probability of use and fixed collocation [21].

3.1.1. Functional Word Diagnosis

Chinese functional words are very rich, which can be divided into adverbs, prepositions, conjunctions, auxiliary words, interjections and onomatopoeia. Except adverbs, they are generally not used as phrases or sentences. They have no real meaning, but have grammatical or functional meaning. Functional words do not exist in English language in a sense, but are reflected in the change of tense and sentence pattern in grammar. In the process of translation, we must make flexible transformation according to the context and the grammatical rules of the target language. To evaluate the machine translation of Chinese functional words is to test whether the machine translation system accurately expresses the grammatical or functional meaning of functional words. All the functional words in the Chinese corpus are labeled manually or automatically. The corpus retrieval tool is used to list all the translation examples of the functional words. Through text observation and language analysis, the translation results of the functional words are scored according to the completion. The score of completion degree is set as [0] and [1], that is, the score of completion (correct translation, accurate meaning) is 1, and the score of unfinished (wrong / untranslated, no / reverse meaning) is 0. Taking the *Selected Works of Xiaoping Deng* (Volume 3) as the test corpus, the translation results of Youdao translation system are taken as the evaluation objects, and the scoring results are shown in Table 1.

Table 1. Results of some functional words in Youdao translation.

Serial number	Functional words with context	Youdao translation	Completion	Completion value
1	(以)党治国	Ruling the country by the Party	The translated words are correct and the meaning is accurate	1
2	因党(而)骄	Party members are "proud of the Party".	Functional words are not translated, and the meaning is opposite	0
3	(有的)变是好事，问题是变什么。	Some of them are good, and the question is what should be changed.	The translated words are correct and the meaning is accurate	1
4	(不)干预行(吗)?	Could we afford not to intervene?	The translated words are correct and the meaning is accurate	1
5	(再)一个问题是军队的纪律(很)差。	Another problem is discipline in the army.	(再) the translated words are correct and the meaning is accurate; (很) the untranslated words have no meaning	1/0

The majority of Chinese functional words show the adaptive transformation of semantic relevance and grammatical rules in English language. In the process of testing, it is necessary to combine the contextual information to judge and deduce the usage standard of the functional words in English language. Therefore, to be familiar with and master the rules of the use of functional words in Chinese and English is the basic condition to evaluate the translation quality.

3.1.2. Word Collocation Diagnosis

Word collocation is a way of meaning expression at the word level [22], which varies with different languages and may lose its original characteristics in the process of translation [21]. There are many factors that affect words and word combination, such as semantics, context, language habits, style, register, etc. in natural language processing, and a variety of factors are introduced. Based on rules, statistics and machine learning methods, the computer can reduce the error rate of word collocation as much as possible in the process of translation. As stated in modern Chinese collocation [23], it is limited by nationality, times, locality, diversity, subjectivity and other factors. The combination methods include subject predicate, verb object, definite middle, adverbial middle, prepositional object and so on. However, English collocations are generally divided into general collocations, rhetorical collocations, professional collocations and conventionalized collocations [24]. According to the collocation intensity, they can be divided into three types: partially restricted collocations, highly restricted collocations and completely restricted collocations [25]. Based on the basic concepts of Chinese English collocation, the evaluation principles of machine translation results of collocation can be divided into three categories: natural principle, conventional principle and creative principle. The principle of nature mainly investigates whether the word form combination conforms to the collocation habits and ways of

the target language; the principle of conventionality, that is, the association and co-occurrence of word form, is based on the conventional rules and norms; the principle of creation is the subjective collocation, but it does not violate the idiomatic and conventionality of word form combination. For example:

Original: 我们不仅要武装战士的手足，尤其重要的是武装战士的头脑

Reference translation: Our soldiers should not only have weapons in their hands but, more important, they should also be armed mentally.

Youdao translation: We should not only arm the arms and feet of soldiers, but also the brains of armed soldiers.

The above example is the collocation form of "V + n" and has semantic motivation. The morphological combination of "armed hands and feet" is a specific concept, which means soldiers hold weapons, while "armed mind" is an abstract concept, which means soldiers keep rational and fresh mind. Based on the standard translation, Youdao translation is too literal and does not conform to the target language norms, which violates the natural principle of collocation.

Original: 这对于军队战斗力的提高，有非常重大的意义。

Reference translation: This is most important to the enhancement of the troops' combat effectiveness.

Youdao translation: This is of great significance to the improvement of the combat effectiveness of the army.

In Chinese, the phrase "great significance" is a definite middle phrase structure, with modifying and modified elements. In English, the fixed collocation "be important / of great significance" can be used to express the importance of something. Therefore, Youdao translation conforms to the conventional principle of collocation in this example.

Original:允许战士家属到营地看望，并给家属以很好的安慰和招待。

Reference translation: We should also allow their family members to visit them in the army barracks, show solicitude for them and entertain them.

Youdao translation: The family members of the soldiers are allowed to visit the camp and give them good comfort and hospitality.

The phrase "allow someone to do something" is a verb object structure. The reference translation adopts the common collocation "allow someone to do something" in English, while Youdao translation adopts the passive voice of "allow someone to do something" without violating the natural and conventional principles of collocation, highlighting the object of action of the verb, which is in line with the creative principle of collocation.

3.2. Sentence Level Evaluation

The evaluation of the quality of machine translation from the sentence level mainly focuses on the diagnosis of consistency of the average sentence length and the word order of interrogative sentences. The average sentence length of standard translation and machine translation is calculated by Python programming language. The larger the value is, the more complex the sentence is. Chinese and English both include declarative sentence, interrogative sentence, exclamatory sentence, imperative sentence and inverted sentence. The significant difference is that interrogative sentence in English can be divided into special interrogative sentence and general interrogative sentence. In written language, special interrogatives such as "where", "what", "why", "how", "who" and "when" are usually placed in the middle or end of Chinese sentences instead of the beginning in English sentences. General questions in English sentences need be-verbs and modal verbs to complete. From the consistency of word order conversion of Chinese and English interrogative sentences, we can judge and evaluate the translation results of machine translation system at the sentence level.

3.2.1. Average Sentence Length Diagnosis

The average sentence length can be used to measure the difficulty of sentences in the text, and the average sentence length is directly proportional to the complexity of sentences [26]. The average sentence length is usually calculated without punctuation and other meaningless sentence elements. Chinese sentences are calculated by the number of "*Zi* (字)", while English sentences are calculated by the number of "*Ci* (词)". By comparing the average sentence length between machine translation and reference translation, the degree of sentence complexity handled by machine translation system can be reflected, and the adequacy of sentence translation in machine translation can be judged (as shown in Table 2).

Table 2. Average sentence length of machine translation "Selected Works of Xiaoping Deng".

The type \ The scope	Volume 01	Volume 02	Volume 03	The three volumes
Reference translation	23.03	20.10	17.90	19.24
Baidu translation	21.46	18.01	16.24	18.92
Youdao translation	20.43	17.28	15.41	17.53

The reference translation of the first and second volumes of *Selected Works of Xiaoping Deng* is taken from the second edition of Foreign Language Press published in 1995, and the reference translation of the third volume is from the first edition in 1994. It can be seen from the table that the sentence complexity of the translation of Volume 1 is high. The average sentence lengths of Baidu translation and Youdao translation are 21.46 and 20.43 respectively, which are 1.57 words and 2.6 words different from the reference translation. The average sentence length of the reference translation of Volume 2 is 20.10, which is 2.09 words and 2.82 words longer than that of Baidu translation and Youdao translation. The average sentence length of the reference translation of Volume 3 is 17.90, which is 1.66 words and 2.49 words longer than that of

Baidu translation and Youdao translation. Considering the whole text, the average sentence length of Baidu translation is closer to that of reference translation, with a difference of only 0.32 words. Therefore, for sentence level processing, Baidu translation can better reflect the sufficiency and integrity of sentence content and information than Youdao translation.

3.2.2. Word Order Diagnosis of Interrogative Sentences

The declarative sentences in Chinese and English belong to the same subject predicate object structure, while the order of interrogative sentences is different. Except for rhetorical questions, interrogative words, auxiliary verbs and modal verbs in English interrogative sentences are usually placed at the beginning of sentences. To judge the result of machine translation of English interrogative sentences is to evaluate whether the use of special interrogative words such as *what, when, why, where, who, how*, be-verbs such as *is, am, are* and modal verbs such as *can, may, shall, will* and their positions in sentences are correct or not. For example, by using Python programming language software to search the sentences with question marks in the full text of *selected works of Xiaoping Deng*, a total of 996 sentences were obtained, of which 23% were interrogative sentences in Volume 1, 44% were interrogative sentences in Volume 2, and 33% were interrogative sentences in volume 3. Statistics of special interrogative words (what, when, why, where, who, how) appear at the beginning of the sentences are shown in Figure 1.
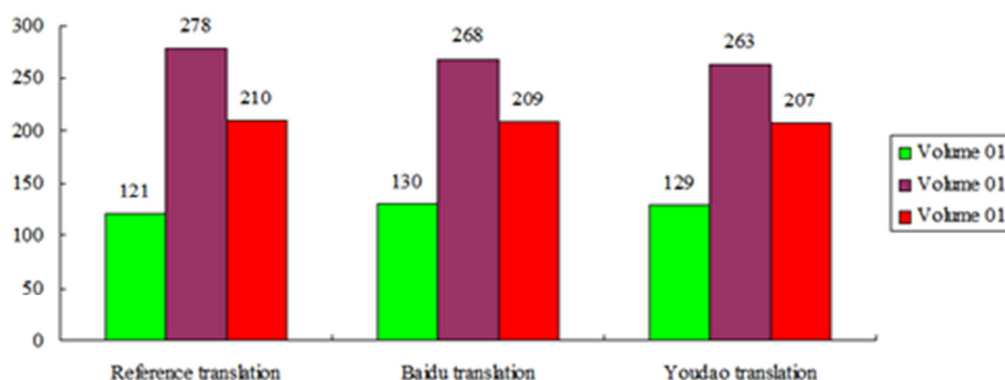

Figure 1. Frequency statistics of special interrogative words at the beginning of sentences.

As can be seen from the figure above, there are 609 special questions in the reference translation, accounting for 61.14% of the total number of questions; 607 special questions in Baidu translation, accounting for 60.94% of the total number of questions; 599 special questions in Youdao translation, accounting for 60.14% of the total number of questions. In terms of the frequency of using special interrogative words at the beginning of sentences, Baidu translation is closer to the reference translation, with a difference of only 0.2%, while Youdao translation is different from the reference translation, with a difference of 1%. Comparing the machine translation with the reference translation, the closer the proportion of the interrogative sentences with the characteristic words at the beginning of the sentence is, the higher the probability of the correct position of the language unit in the sentence is.

3.3. Semantic Level Evaluation

Semantic analysis means that the machine translation system obtains formal language related background knowledge, which is a very complex and important link after lexical analysis and syntactic analysis. Chinese and English have different ways of expressing meaning. The former expresses the grammatical function and logical relationship of sentences through the meaning of language units, while the latter obtains the logical meaning of sentences through the connection of structure and form of language units. Therefore, the Chinese-English machine translation system needs to analyze the grammatical function and meaning of the sentence components in the source language in order to determine the complete semantic information of the sentence. By comparing the text similarity between the machine translation and the reference translation, the understanding ability of the machine translation system is judged. The closer the similarity value is, the stronger the understanding ability of the machine translation system is, which means that the semantic information of the machine translation is more complete.

3.3.1. Text Similarity Diagnosis

In the process of translation, in order to convey complete semantic information, the translator (or machine translation system) should try to maintain the similarity between the target language text and the source language text. Taking the reference translation as a standard, the vector space of the words in the translated text is established by using cosine similarity measure method, and then the cosine value of the angle between the two vectors is calculated through the cosine function formula. The larger the value is, the higher the similarity between the two texts is. The specific operation is as follows:

source text: 政府机关支部与一般农村支部的对象不同，它还有其特殊的任务，而且要有比较健全的支部委员会的领导，才能实现这些特殊的任务。

reference translation: Unlike those in rural areas, Party branches in the government have their own special tasks, which require sound leadership from the Party branch committee.

Baidu translation: Different from the general rural branch,

the branch of government organs has its own special tasks, which can only be accomplished under the leadership of a relatively sound branch committee.

Youdao translation: Different from the general rural party branch, the branch of a government organ has its own special tasks, which can only be realized by the leadership of a relatively sound party branch committee.

To construct the dictionary of reference translation and Baidu translation d_1:

d_1=['those', 'tasks', 'only', 'areas', 'can', 'its', 'under', 'has', 'accomplished', 'relatively', 'government', 'require', 'have', 'party', 'of', 'different', 'which', 'in', 'rural', 'a', 'branches', 'the', 'from', 'sound', 'own', 'special', 'general', 'branch', 'committee', 'organs', 'leadership', 'unlike', 'be', 'their']

The frequency of each word in the reference translation and Baidu translation is calculated and its vector representation is constructed as follows:

a) reference translation: t=[1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 2, 0, 0, 1, 2, 1, 0, 1, 2, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1]

b) Baidu translation: t_1=[0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 2, 1, 1, 0, 1, 1, 0, 3, 1, 1, 1, 1, 1, 3, 1, 1, 1, 0, 1, 0]

The cosine similarity of t and t_1 is calculated by cosine function:

$$\cos(\theta) = \frac{\sum_{i=1}^{n}(x_i \times y_i)}{\sqrt{\sum_{i=1}^{n}(x_i)^2} + \sqrt{\sum_{i=1}^{n}(y_i)^2}}$$

$$\cos(t, t_1) = 0.52295779$$

Similarly,

To construct the dictionary of reference translation and Youdao translation d_2:

d_2=['those', 'tasks', 'only', 'areas', 'can', 'its', 'has', 'relatively', 'government', 'realized', 'require', 'have', 'party', 'of', 'different', 'which', 'in', 'rural', 'a', 'branches', 'the', 'from', 'sound', 'by', 'own', 'special', 'general', 'organ', 'branch', 'committee', 'leadership', 'unlike', 'be', 'their']

The frequency of each word in the reference translation and Youdao translation is calculated and its vector representation is constructed as follows:

a) reference translation: t=[1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 2, 0, 0, 1, 2, 1, 0, 1, 2, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1]

b) Youdao translation: t_2=[0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 2, 2, 1, 1, 0, 1, 2, 0, 3, 1, 1, 1, 1, 1, 1, 3, 1, 1, 0, 1, 0]

The cosine similarity of t and t_2 is calculated by cosine function:

$$\cos(\theta) = \frac{\sum_{i=1}^{n}(x_i \times y_i)}{\sqrt{\sum_{i=1}^{n}(x_i)^2} + \sqrt{\sum_{i=1}^{n}(y_i)^2}}$$

$$\cos(t, t_2) = 0.58800649$$

From the above results, it can be seen that the similarity between Youdao translation and reference translation is greater than that between Baidu translation and reference translation. Therefore, it can be judged that the semantic translation of Youdao translation is closer to the theme, thought and emotion of the original sentence. Similarly, the similarity of a text can also be calculated,

The similarity values of Baidu translation and reference translation are 0.98770573, 0.98757994 and 0.98283705 respectively, and the similarity values of Youdao translation and reference translation are 0.98540654, 0.98672976 and 0.98432985 respectively. By comparing the similarity results, we can get the semantic translation effect of different texts in Baidu translation and Youdao translation. For example, the semantic effect of Baidu translation in Volume 1 and Volume 2 is better than that of Youdao translation in volume 3.

3.3.2. Key Words Diagnosis

Key words are subject words that summarize and convey the content and information of the text, also known as subject words. Through the corpus retrieval software, all the application examples of key node words in the corpus are listed, and the frequency of the words in the text is counted. By comparing the keyword index sequences of the machine translation and the reference translation, we can judge whether the machine translation is variable translation or stable translation according to the degree of change. Secondly, we count the frequency of keywords used in the original text, in the machine translation and in the reference translation, and regard the frequency of keywords used in the original text as F1, the frequency of keywords used in the machine translation as F2, and the frequency of keywords used in the reference translation as F3. The hypothesis is that the closer the F2 value is to the F1 value, the higher the fidelity of the machine translation, and the closer the F2 value is to the F3 value, the better the naturalness of the machine translation. For example, in the first article of the *selected works of Xiaoping Deng* (Volume I), *Mobilization of New Soldiers and Political Work for New Soldiers* (January 12, 1938), the key node word is "*bing* (兵)", and the official term is "recruit". For the sake of vocabulary rules in English, it is searched by word roots, and the results are as follows:

In the source text, "*bing* (兵)" appears 22 times in such collocations literally as "mobilize new soldiers", "new soldiers politics", "soldier corps", "being a soldier", "recruit soldiers", "military service", "officers and men", "recruiting system", and "old soldiers", etc.

"recruit" appears 24 times in the reference translation. The collocations are: mobile new recruits, enlist new recruits, recruiting soldiers, recruitment efforts, methods of recruitment, recruitment methods, carry out recruitment;

"recruit" appears 25 times in Youdao translation, with the following collocations: mobilization of recruits, mobilizing recruits, the way we recruit, recruiting troups, recruitment and redeployment work, way of recruitment, recruit new solvers, recruit mobilization.

From the above analysis, Youdao translation is generally consistent with the reference translation, which belongs to stable translation. From the frequency of key words, there are 22 collocations of "*bing* (兵)" in the original text (that is, F1=22), while "recruit" appears 25 times in Youdao translation (that is, F2=25) and 24 times in reference translation (that is, F3=24), so F2 value is closer to F3 value.

Thus, compared with faithfulness, Youdao translation performs better in target language understanding, expression and fluency.

3.4. Pragmatic Evaluation

It is worth trying to introduce the theoretical knowledge of pragmatics into computer technology. Pragmatics is the study of language use and its laws. The implied meaning and communicative effect of language use can only be reflected in the specific language environment. Therefore, understanding the logical relevance of the context information of the source language can make the target language correctly reflect the real meaning and intention of the original text. Text sentiment analysis reflects the subjective information of a text and obtains the speaker's emotional state and communication intention by mining the text content. To compare the affective orientation value (E1) of the machine translation with the affective orientation value (E2) of the reference translation, the closer the E1 value is to the E2 value, the more accurate the machine translation is in understanding the original context, that is, the more sufficient the machine translation is. For example:

Source Text: 当前的战局，是处于暂时的局部的失利的境况，决不是抗日自卫战争的最后失败。

Reference transaltion: Currently we are suffering a temporary and partial setback in our defensive war against Japan, but this is not final defeat.

(Emotional polarity: positive; Positive probability: 0.99920017; Negative probability: 0.0007998973)

Baidu translation: The current situation of war is a temporary local defeat, and it is by no means the final defeat of the war of self-defence against Japan.

(Emotional polarity: positive; Positive probability: 0.9954066; Negative probability: 0.0045934017)

Youdao translation: The present situation is one of temporary and partial defeat and is by no means the final defeat of the war of resistance and self-defence.

(Emotional polarity: positive; Positive probability: 0.9972767; Negative probability: 0.0027232843)

From the above examples, the words "temporary", "partial" and "by no means" in the original text indicate that the speaker's attitude is positive. Therefore, the emotional polarity of the machine translation conforms to the subjective emotional tendency of the original text. The value of emotional polarity is 1, with 0.5 as the cut-off point, greater than is positive emotion and less than is negative emotion. The positive and negative affective strength of the text is reflected by positive probability and negative probability. The former reflects the positive affective strength of the influencing words in the text, while the latter reflects the negative affective strength. The emotional polarity of the reference translation is positive, and the emotional strength is very close to 1. Baidu translation and Youdao translation both belong to positive emotion tendency, and the emotional strength is 0.9954066 and 0.9972767 respectively, which is 0.00379357 and 0.00192347 different from the positive emotion strength of the reference

translation. Therefore, in this sentence, Youdao translation has more accurate understanding and processing of the text context, and the adequacy of the translation is higher.

4. Conclusion

In the era of Digital Humanities, the combination of digital technology and translation studies is reflected in the use of various information tools and means to provide corpus or data for the explanation of translation phenomena. Chinese-English machine translation quality assessment model uses a variety of technical means to evaluate the effect of machine translation system, and takes linguistic knowledge and translation criteria as the theoretical guidance of quality assessment, so as to provide a practical and effective machine translation quality assessment model for researchers and translators. However, there are still some limitations in this model, which need to be further studied.

References

[1]   Busa R. The annals of humanities computing: The Index Thomisticus [J]. Computers and the Humanities, 1980, 14 (2): 83-90.

[2]   Yang Xianyu, Rongguang Yang. Summary of Chinese English translation [M]. Tianjin University Press, 2015.

[3]   Feng Zhiwei. Formal model of natural language processing [M]. China University of science and Technology Press, 2010.

[4]   House, J. A Model for Assessing Translation Quality [J]. Meta, 1977, 22 (2): 103109.

[5]   Toury, G. Descriptive Translation Studies and Beyond [M]. Amsterdam: John Benjamins, 1995.

[6]   Nida, E. A. Toward a Science of Translating [M]. Leiden: E. J. Brill, 1964.

[7]   He Sanning. Research on translation quality assessment model [M]. Central Compilation and Translation Press, 2015.

[8]   Ba Xiwei. Reflections on the instrumental nature and translation criteria of translation [J]. China Science and technology translation, 2016, 29 (04): 48-50.

[9]   Liu Junping, Qin Jianghua. Selected readings of western translation theories [M]. Wuhan: Wuhan University Press, 2012.

[10]   Xiao Kunxue, Lu Daofu. College English general translation course [M]. Jinan University Press, 2012.

[11]   Sun Yiqun. Research on machine translation quality evaluation method based on questionnaire and data analysis [J]. China foreign language research, 2018: 72-77.

[12]   Sun Xiao, Zhu Conghui, Zhao Tiejun. Machine translation quality estimation algorithm based on translation knowledge [J]. Intelligent computer and application, 2019: 271-275.

[13]   Yang Zhongcheng. Machine translation quality evaluation based on pretraining language model [D]. Beijing Jiaotong University, 2019.

[14] Hou Qi. Research on machine translation quality evaluation method based on knowledge transfer [D]. Nanjing University, 2019.

[15] Wu Ping. Research on AHP fuzzy evaluation method of machine translation quality in English translation [J]. Information technology, 2019: 107-111.

[16] Li Peiyun, Zhai Yujin, Xiang Qingyu, et al. Translation quality estimation method for sentence level neural machine translation based on subwords [J]. Journal of Xiamen University: Natural Science Edition, 2020: 159-166.

[17] Li Peiyun, Li Maoxi, Qiu Bailian, et al. Research on translation quality estimation method based on word vector of best context [J]. Chinese Journal of information, 2020: 60-67.

[18] Lu Jinliang, Zhang Jiajun. Translation quality estimation method based on multilanguage pretraining language model [J]. Journal of Xiamen University (NATURAL SCIENCE EDITION), 2020: 151-158.

[19] Zhang Xiaojun. Computational linguistics [M]. Shanxi Normal University Press, 2011.

[20] Firth, J. R. Modes of Meaning [A]: (In) J. R. Firth, Paper in Linguistics 19341951 [M]. London: Oxford University Press, 1957.

[21] Wu Zhaolu, Jia Fei Sheng Er, Lin Jun Xiang. Chinese Studies (Volume 11) [M]. Jinan press, 2008.

[22] Zhang Jidong. A study on the heterogeneity of collocations in EST [M]. Shanghai Jiaotong University Press, 2013.

[23] Li Yude. Collocation of modern Chinese [M]. Commercial Press International Limited, 1998.

[24] Wei Naixing. Definition and research system of collocation [M]. Shanghai Jiaotong University Press, 2002.

[25] Wei Naixing. Essentials of lexicology [M]. Shanghai Foreign Language Education Press, 2011.

[26] Huang Libo. Corpus based study of translation style [M]. Shanghai Jiaotong University Press, 2014.