Application of Corpus Analysis Software and Error Analysis Theory on Non-English Majors' Chinese-English Translation Features

Can Shu

School of Foreign Language, Wuhan Business University, Dongfeng Boulevard, Wuhan, China

Abstract: This research focuses on non-English major students' Chinese-English (C-E) translation features. It aims at annotating and summarizing the features and error types in students' C-E translation by applying to corpus analysis software Readability Analyzer. This study first collects students' C-E translation texts from a computer-based exam, then forms a self-constructed corpus and a reference corpus. The translation materials in the target corpus and the reference corpus are first analyzed and compared by Readability Analyzer to figure out the features of students C-E translation texts from the perspectives of sentence length, word length, percentage of difficult words, Flesch Reading Ease, Dale-Chall Score, Fry Readability Grade Level. After that, different types of errors in students C-E translation are labeled and analyzed based on statistical analysis and Error Analysis (EA) theory. According to Readability Analyzer, students C-E translation texts have the following features. From passage statistics, it shows that translation texts that get lower scores tend to use shorter sentences and easier words. In addition, from readability scores, Flesch scores, Dale-Chall Scores shows and Fry Readability Grade Level are taken into consideration. Flesch scores shows that compared with the reference corpus, students' translation which get lower scores are harder to understand since there are more language errors than higher-score translation. Dale-Chall Scores indicate that translation materials which score 8-9 are very close in the difficulty of word-using with the reference corpus. Fry Readability Grade Level shows that texts of lower scores are simpler and more accurate than texts of higher scores. According to Error Analysis (EA) Theory, students' errors are further classified and discussed on lexical level, sentence level and textual level. The conclusions of this research are expected to make contributions to the improvement of C-E translation teaching and help college students, especially non-English majors, to enhance their C-E translation competence.

Keywords: Chinese-English translation; corpus analysis; Error Analysis Theory

1. Readability Analyzer

Readability Analyzer is a corpus analysis software programmed and designed by Jia Yunlong and Xu Jiajin from National Research Center of Foreign Language Education in Beijing Foreign Studies University. This software can estimate the readability of a passage by providing passage statistics and readability scores.

From all the metrics above, this research chooses to collect and analyze Flesch Reading Ease, Dale-Chall Score and Fry Readability Grade Level. As the Gunning Fog is similar to Flesch score, this research only compares the Flesch score of the translation texts. And SMOG score requires passage to be at least 30 sentences long, and the translation texts are all short passage less than 100 words, so SMOG grade is not included in this research. The Flesch-Kincaid Grade Level is not so relevant to this research purposes, so it is also excluded from this research.

2. Error Analysis Theory

In 1967, Pit Corder first gave the definition of "error" in his article The Significance of Learner's Errors, he emphasized the important meaning of recognizing and analyzing errors in second language teaching and learning. He also put forward the basic theoretical framework of error analysis theory. In error

analysis theory, the basic procedure is to first collect the research material, then to make a distinction between errors and mistakes, because errors which occur because of learner's lack of language competence are considered worth analyzing, however, mistakes which appear because of students' carelessness or negligence are not included in an error analysis research. Carl James (1998) claimed that "If the learner is inclined and able to correct a fault in his and her output, it is assumed that the form he or she selected was not the one intended, and we shall say that the fault is a mistake. If, on the other hand, the learner is unable or in any way disinclined to make the correction, we assume that the form the learner used was the one intended, and that is an error."

For this research, the errors in students' C-E translation works are going to be divided into lexical level, sentence level and textual level. On lexical level, errors in word spelling, inappropriate use of words will be included. On sentence level, errors are mainly about the incorrect organization of sentence structures, errors in grammar and missing translation. On textual level, errors in coherent and the logical relationship among sentences need to be considered.

3. Methodology

3.1 Research Questions

This research aims at discussing and finding out the answers to the following questions.

(1) Do students' C-E translation texts have any apparent features on lexical, sentence and textual levels?

(2) How to define and distinguish different types of errors appear in the corpus?

(3) What types of errors do non-English students have when they do C-E translation?

3.2 Research Subjects

This research has selected 60 pieces of C-E translation texts from 60 students in X University. The students are all second-year students from non-English majors. Because this university has divided all non-English majors into A, B and C level classes according to their English test scores in Chinese Entrance Exams for Universities, these students are all from B level classes, thus can represent the average English level of all the non-English major students in the university. The C-E translation they accomplished is from their final exam for College English II course, which is a compulsory course that all non-English students are required to take. So, it can be ensured that the students were trying their best to accomplish the translation task.

3.3 Procedures

The students are required to translate a 150-word Chinese material into English on computers.

Students' translation will be first scored according to CET-4 translation scoring standard, then researchers will try to annotate the translation material and classify the error types according to this research questions and purposes.

Then, students' translation texts will be further analysed by Readability Analyzer. Then the errors in C-E translation will be analysed from three levels--word, sentence and text levels.

At last, C-E translation features of non-English majors will be summarized according to theprevious data collection and analysis.

4. Data Analysis

4.1 Readability Analyzer

Table 1. Readability Analysis

	Scores	3-4	5-6	7-8	Reference corpus
Passage Statistics	Words Per Sentence	14.8	16.2	17.1	20.3
	Characters Per Word	5.1	5.2	5.3	4.9
	Percentage of Difficult Words	15.5%	15.9%	18.7%	18.1%
Readability Scores	Flesch Reading Ease	49.1	46.5	42.1	45.6

Table 1 shows the readability of a text from several perspectives. In order to analyse the readability of students' translation in a more detailed way, this research compare the statistics according to different fractional segments, and also build a reference corpus which includes the C-E translation exercise in students' textbooks.

From readability scores, there are several parameters which need explanation in detail.

First, according to Table 1, translation texts score 3-4 get higher Flesch scores, which means that these texts are more difficult to read and understand than others, which probably means that there are more language mistakes, therefore more difficult to understand.

Second, Dale-Chall is one of the most accurate readability metrics. And according to Table 1, translation texts score 3-4 get 8.4 Dale-Chall score, texts score 5-6 get 8.7 Dale-Chall score, texts score 7-8 have the same Dale-Chall score as the reference corpus, they both have 9.1 Dale-Chall score.

As for Fry Readability Grade Level. Higher level means less readable. Translation texts of 3-4 scores have the same Fry Readability Grade Level as the reference corpus, and texts in other score zones get higher Fry Readability Grade Level. So, texts of lower scores are simpler and more accurate than texts of higher scores maybe because of the choice of simpler words and sentence structures.

4.2 Analysis on Lexical Level

Students' C-E translation errors most frequently occur in lexical level. Therefore, it is necessary to further divide errors into more detailed parts. From the translation texts, we can classify two sub-categories in lexical level, that is, the misspelling of words and the inappropriate use of words. According to Table 2, both two types of errors appear frequently, indicating that vocabulary accumulation is still the most urgent thing for non-English major students to do if they want to improve their translation competence.

Table 2. Translation Errors in Lexical Level

	error type	number of errors
Lexical level	misspelling	204
	inappropriate use of words	244

(1) Accounding to a survy obvious that many students read books less than 1 hour,...

(2) Specilest pointed,the fragmented, utilitarian read receds the students'ability of considering and anylax. They considered ,syudents should read humanity works to prove themselves' level.

(3) In this information dynast,college's reading habit became more utilitarian.

Errors due to inappropriate use of words appear most frequently (244 times).

Misspelling errors happen due to two reasons--students' vocabulary inadequacy and failing to understand English grammar in word level. For example, "accounding" and "survy" in sentence(1) belong to the first reason. Students cannot remember how to spell some words, therefore spell these words in an incorrect way. In sentence (2), "specilest" and "anylax" are also considered to be this type.

4.3 Analysis on Sentence Level

Errors in sentence level can be sub-divided into three kinds. Errors in sentence structure outweigh the other types of error greatly, because when students try to write longer and more complicated sentences, they are likely to miss some grammatical parts of the sentence.

(4) A survey shows that many university students reading is not enough an hour every day, but using time online is more than four hours.

(5) Nowadays, the younger likes reading some book that more useful

(6) They do that students should more read humanity works to develop themselves

For example, in sentence (4), "students reading is not enough an hour" is not an appropriate expression, it should be the time they spend on reading, but not reading itself is less than an hour. What's more, in the sentence "but using time online is more than four hours", the subject is incorrect, and the predicate is absent. Therefore, the sentence can be improved as "for many university students, the time they spend on reading is less than an hour every day, but the time they spend online is more than four hours daily."

Grammatical errors in sentence level can be observed in sentence (5). Errors on this level mainly indicate subject-predicate inconsistency, or incorrect collocation between verbs and prepositions. In sentence (5), "the younger likes" should be "the young like". Besides, the+adjective structure indicates a group of people who share the same characteristic, so the predicate should be "like" but not "likes". This inconsistency between subject and predicate is considered a grammatical error in sentence level. Also, in sentence (5), "some book" should be revised as "some books".

Missing translation in sentence level means that students neglect part of a sentence or even the whole sentence, as is shown in sentence (6).

4.4 Analysis on Textual level

(7) In this information exploded age, colloge students' reading habits become utilitarian. A reach outcome shows that,many colloge students read less than one hour, but their online time over four hours. Now the young person like to read more practical books, such as exam directions and career compuse. The expert point out,fragmented, profiled reading receded students' thinking talent. They said one more time, students should read more humanity works to improve their tastes.

In the above paragraph, there are two errors belong to textual level--"The expert point out, fragmented, profiled reading receded students' thinking talent" and "They said one more time, students should read more humanity works to improve their tastes". These two sentences are considered translation errors in textual level because they both use past tense. When taking the translation text as a whole, these two sentences are not very natural in logic, therefore, errors like these are considered textual errors in the C-E translation.

5. Conclusion

This research focuses on non-English major students' C-E translation. According to the analysis, students' difficulties in C-E translation can be summarized as followed.

(1) Students' insufficient vocabulary is the main factor that causes their difficulty in C-E translation. They neither remember the correct spelling of certain words or know the exact meaning of related words, which cause the errors in lexical level, and also the missing translation in sentence level.

(2) In sentence level, the majority of students in the research don't have missing translation, which indicate that when dealing with C-E translation, they make an effort to use their language knowledge to translate. But when they are writing long sentences, errors in sentence structure appear frequently. This shows that students are lack of sense for even the simple basic English grammar.

(3) In textual level, the errors occur simply because of students' lack of grammar knowledge, or maybe they just need to read through the whole text when they finish their translation and pay more attention to tense and personal pronouns.

References

[1] C. James. "Errors in language Learning and Use: Exploring Error Analysis". Beijing: Foreign Language Teaching and Research Press, 2001, p78.

[2] S. P. Corder, "The significance of learners' errors". IRAL, 5, 1967, p161-170.

[3] S. P. Corder, "Describing the language learner's language". CILT Reports and papers NO.6. CILT.1971.

[4] S. P. Corder, Error Analysis. In Allen,J & Corder, S. P. (eds.), "The Edinburgh Course in Applied Linguistics", Vol 3. Oxford: Oxford University Press, 1974.

[5] H. Dulay, M. Burt, & S. Krashen. "Language Two". New York: Oxford University Press.1982.

[6] J. Richards, "Error analysis". London: Longman. 1974.