A Parallel Corpus-Based Study of the Difference in Dependency Distance Between English and Chinese in C-E Translation

LI Liang
Wuhan University of Technology, Wuhan, China

Based upon four self-built parallel corpora, the paper presents itself as a comparative study of the dependency distance between Chinese ST and English TT, between the works of the same style and between the works of different styles. The result shows that: (1) The dependency distance value (the DD value) of Chinese is much larger than of English; (2) English is a governor-initial language, while Chinese tends to be more governor-final; and (3) many factors, for example, the position of governors and dependents, the frequency of the connections between sub-clauses, difference in inflectional explicitation of C-E translations, concise purpose as well as translative styles have led to the difference in dependency distance between Chinese and English.

Keywords: parallel corpus, contrastive study, dependency distance, dependency direction

Introduction

Dependency distance refers to the linear distance from a governor to a dependent, having a syntactic relation within a clause, that is, “the distance between words and their parents, measured in terms of intervening words” (Hudson, 1995). In the theory of dependency grammar, the conception was first brought in by Heinger, Strcker, and Wimmer (1980). Under the framework of dependency grammar, for any dependency relation between the words W1 and W2, if W1 governs W2, the dependency distance value can be defined as the difference a-b. And if the linear order of the two is that W1 precedes W2, then the DD value should be a negative number and the dependent direction is said to be governor-initial, thus exocentric; if Wb governs Wa in the same linear order that Wa precedes Wb, then the DD value should be negative, and the dependency direction is head-final, thus endocentric.

Regarded as a metric reflecting the comprehending difficulty (Liu, 2008), dependency distance is of great benefits to linguistic studies. The processing as well as analyzing of a clause is to translate a sequence of linear signs into a dependency relation, that is to say, a word can be removed from human’s short memory if and only if it is connected with other words to form a dependency relation.

At the same time, dependency distance may reflect the direction of a dependency, which serves as an indicator for word-order typology (Liu, 2010). It can be employed as an important tool for identifying the typological characteristics of a language as whether it is head-final or dependent-final.

However, although there have been many statistics and analyses for the dependency distance of English and Chinese, most of them tend to be specific on a certain language or refined to the comparison between the same style of texts, not only leaving scant studies conducted in terms of the comparison in the dependency...
distance between the two languages together, but also in terms of the comparison between texts with different forms of style.

With the tools of SPSS, AntConc, Trados, and Corpus Word Parser, and by means of four parallel corpora with different genres, including *Report on the Work of the Government 2019*, *China’s National Defense in the New Era*, and *President Xi calls for targeted poverty alleviation*, this paper is designed to make a retrieval of dependency distance along with direction and present a comparative study of the dependency distance in both Chinese and English to get insights into the comparative difference in the dependency distance of the two languages and the reasons accounting for the disparity.

**The Corpus-Based Studies of Dependency Distance**

Corpus translation studies have been a newly-arising branch of subject for recent decades in the field of translation. Compared with traditional translation researches, corpus translation puts more emphasis on empirical studies, the integration of theory analyses and data statistics (Huang, 2012). Based on the conceptions and methods of corpus linguistics, guided by linguistics and translation theories, instrumented by data statistics and possibility analyses, targeting naturally-occurring utterances in bilingual corpora, corpus translation studies, through descriptive translation studies, is linked with corpus linguistics to conduct researches on synchronic and diachronic studies of translation, to make systematic analyses in translation nature, process and phenomena. “Corpus translation has its own research fields, such as linguistic features and styles of translation texts and their translators” (Hu, 2012, pp. 59-68).

According to Eugene A. Nida (1982), from the perspective of linguistics, the key difference in Chinese and English is that the former is a paratactic language while the latter is hypotactic. In English, a verb tends to be transformed into an abstract noun to convey abstract concepts, whereas in Chinese, verbs are used to convey specific acts. In this sense, English is of more conciseness, thus usually an English TL is less lengthy than its original Chinese ST. In this sense, is it justified to assume that it is the built-in syntactic characteristics of conciseness and compactness that make the value of dependency distance in English lower than in Chinese? And considering the fact that although the specific studies of dependency distance values of Chinese and English have been conducted, there are scant researches into comparative studies of both languages and the influence of the role played by different genres in sample texts has received scarce attention, it is necessary to analyze whether different genres can exert effects upon the dependency distance value of texts. While with the intervention of parallel corpora, which has been of great assistance to translations by conducting quantitative research and quantitative analysis in order to reveal the nature of translation as a medium for communication, can be instrumental to researches into the disparities in dependency distance value between texts of the same or different genres and the reasons accounting for them.

**Research Method**

**Research Subject and Research Questions**


**The Setting up of Parallel Corpora**

First, the parallel corpus-based e-texts are stored in the format of txt and the application of translation
memory is used for the alignment of documents; then comes to the syntactic tagging. The whole process adopts the POS-tagging method right after the numbering of the words in each clause. With about 2,700 characters each, roughly equivalent division of the total utterances into two groups is adopted. The POS-tagging work of the second group was in the charge of another three senior students of English major in WUT. With all the tagging work done, the author should finish the proofreading and revision processes, with the major aim to correct and revise obvious tagging mistakes and errors, especially those that would make clauses syntactically and grammatically unacceptable. Meanwhile, attention should also be paid to the utterances preprocessing, such as SBC case of numbers and the standardization of punctuation of Chinese and English.

Results and Discussions

Statistical Processing

First, in the parallel corpora of China’s National Defense in the New Era (2019), Report on the Work of the Government 2019, and the news report President Xi calls for targeted poverty alleviation, the DD value, MDD value, and the number of dependency relations are respectively calculated and listed. Then, of each text, the analysis of the number of positive and negative dependency relations, as well as their average DD value is conducted. At last, the Chi-square tool of SPSS is employed for the inspection of $p$-value.

Statistical Analysis

The comparison of the difference in dependency distance value between Chinese ST and English TT. With the dependency relations of punctuations (bnd, bjd) as dependents removed, and the reference value of the dependency distance of the bjd, as governors, set as 1, the Chinese ST of the excerpt of China’s National Defense in the New Era (2019) has an average DD value of 0.8, while that of its English TT, is 0.34. Then, by means of the Chi-squared distribution, the two average DD values possess a statistically significant difference ($p = 0.00 > 0.05$). The sum total of the Chinese ST dependency relations is 161, and that of its English TT dependency relations is 237. Among those dependency relations, in Chinese ST, the DD value of 113 dependency relations, with the average DD value of 25.97, is greater than 0, accounting for 69.43%; 53 of the relations, accounting for 30.57%, with the average DD value of -21.02, have the 0-below DD value. And in English TT, with the average DD value as 25.24, accounting for 60.75%, 144 dependency relations share the DD value greater than 0; 93 dependency relations have the 0-below DD value, taking up 39.24%, with the average DD value as -23.56. According to the results shown by the chi-squared distribution, there is a significant difference between the Chinese ST and the English TT, not only in the number of the positive relations ($p = 0.017 < 0.05$), but also in the respective average DD values of the positive dependency relations ($p = 0.000 < 0.01$). At the same time, if the dependency direction is not considered, the average MDD value of the Chinese ST is also much greater than that of its English TT, with the former numbered as 1.99, whereas the latter as 2.03.

The result of the Excerpt of China’s National Defense in the New Era (2019) shows that although the p-values of the number of positive dependency relations and of the average MDD do not suggest a significant difference ($p = 0.096 > 0.05, p = 0.090 > 0.05$), there still exist a statistically significant distinctions in the average DD value between the Chinese STs and their English translations, with the former calculated as 0.02, and the latter as -0.40 ($p = 0.00 < 0.01$).

The distinction in the figures shows the tendency that first, from the perspective of language typology, Chinese is deemed as a governor-final language because of the larger number of the dependency relations with
DD values that are positive in the Chinese ST itself than with negative ones and also because the Chinese STs, compared with the English translations, have the greater number of dependency relations sharing positive DD value.

From the perspective of dependency relation, the syntactic structures of Chinese are not only manifested in the aspect that merely showing percentage of the dependents following the governors, with a certain number of dependency relations and a negative DD value, but also in the aspect that the DD value are much greater when the dependency relation is governor-final than when it is governor-initial.

For the two reports, not only the $p$-value of the average DD value ($p = 0.051 > 0.001$) from the aspect of the total number, but also the $p$-value of the average number of positive dependency relations ($p = 0.667 > 0.01$) suggests there is no signs of a statistically significant difference between the excerpt of China’s National Defense in the New Era (2019) and Report on the Work of the Government 2019, despite their statistically significant difference in the average MDD value ($p = 0.00 < 0.01$).

The influencing factors of the difference in the dependency distance and dependency direction. Multi-faced factors contribute to the difference in dependency distance and dependency direction.

First, from the perspective of language typology, typological difference can lead to the distinction in syntactic constructions. Chinese is a language with highly frequent uses of verbs and its prevalent syntactic structures are SV and AdjN. However, different from Chinese, English views nouns as the most-frequent used word class. And one of the most popular syntactic constructions in English is the VO and NPP constructions. One case in point is that complemental pre-modification in Chinese. While in English, prepositional phrases as compliments tend to be put behind their center words, it is more common for the Chinese language to adopt the syntactic construction “X 的 Y” with the Chinese preposition “的” put before nouns as premodifiers and attributes. And if there are quantifiers in the Chinese syntactic construction “X 的 Y” intended to modify a noun word, then farer dependency distance will occur between the noun and the modifying verb.

Second, from the aspect of language characteristics, different characteristics that are formed due to different language families also should be responsible for the disparity of dependency distance. In terms of language classifications, Chinese belongs to the Sino-Tibetan language family in the Asian continent whereas English language is a member of the West Germanic group of the Germanic subfamily of the Indo-European family of languages in the European continent. The disparity in language families consequently causes the difference in language characteristics. From the perspective of syntax, linguistic studies have shown that the major distinction between Chinese and English lies in the distinction between parataxis and hypotaxis, which means, on one hand, English stresses greater importance to syntactic constructions and structures than Chinese, and thus more usages of the dependent or subordinate construction or relationship of clauses without connectives in Chinese than in English. Therefore, in Chinese relatively loose structures of syntax and the uses of dependent or subordinate clause can be more frequently traced than in English, which is the corollary of the greater number of dependency relations and the larger average DD value of these dependency relations between dependent or subordinate clauses in Chinese than in English.

At last, pragmatic factors can also be held accountable for the difference in dependency distance between the two languages. Due to the fact that diverse cultural connotations are shared by Chinese and English, many translators feel it necessary to adopt some translative strategies, such as omission for the purposes of text conciseness and the avoidance of duplication and word redundancy. For example, the translation of the phrase
“亚洲特色安全构架” from the excerpt of China’s National Defense in the New Era (2019) is “Asian security architecture”. The omission is, in essence, an implication of information, realized by the transition from the unknown information and connotations in original texts to the known information and connotations in translated texts in order to make the translated texts more brief, concise, and clear. As a result, the dependency distance of English translated texts is accordingly shorter because of certain translative strategies like omission for language streamlining.

**Conclusion**

In this paper, based on the self-built Chinese-English parallel corpus Report on the Work of the Government 2019, compared with other two self-built Chinese-English parallel corpora, the difference in dependency distance is discussed and analyzed from the angle of comparisons of Chinese-English translations. The results indicate that because of distinct grammatical constructions and cultural connotations caused by the differences in language typology and language families of the two languages, manifested in the aspects of morphology, syntax and pragmatics, translators consciously or unconsciously adopt some translative methods, such as implication and omission for discourse authenticity, conciseness and clarity. On the other hand, the corpus-based comparative studies also shed some lights on the important role of genre in determining the dependency distance. At last, despite the fact that a statistically significant distinction in dependency distance cannot be traced and proven in the comparison of the two political reports, there still exists the difference in the average MDD value, which may be accounted for by the difference in translation styles of translators.

**References**


