

## Translation Research Based on Corpus From Contemporary Linguistics Perspectives

DONG Jinyu<sup>[a],\*</sup>

<sup>[a]</sup>Associate professor. School of Foreign Languages, Northeast Petroleum University, Daqing, China.

Research areas: Linguistics and business English teaching.

\*Corresponding author.

**Supported by** Humanities and Social Sciences in Heilongjiang Province (General project): Study on the Strategies of Spirit of Daqing Going Out (No.13B009).

Received 10 January 2013; accepted 12 April 2014

Published online 26 May 2014

### Abstract

Since 1990s, approaches to translation from contemporary linguistics perspectives have made a number of new progresses. The translation research based on corpus has become the research subject of many scholars. This thesis introduces the theoretical sources of Corpus Translation Studies and also discusses the presumptions of translation language variants and analysis modes based on the interaction of sociolinguistics. At the same time, it combines with other analysis tools, connects the discovery process and the generation process of text together and investigates on the discovery of quantitative research form social, cultural, historical, political and cognitive point of view. This will be the direction of further development on the corpus translation studies in the future.

**Key words:** Linguistics perspective; Corpus; Translation

Dong, J. Y. (2014). Translation Research Based on Corpus From Contemporary Linguistics Perspectives. *Higher Education of Social Science*, 6(3), 45-48. Available from: URL: <http://www.cscanada.net/index.php/hess/article/view/4616>  
DOI: <http://dx.doi.org/10.3968/4616>

### INTRODUCTION

Since 1990s, another eye-catching translation development of linguistics approach studies is the translation study based on corpus. At the beginning of 1990s, the University

of Manchester institute of science and technology set up the first comparative translation corpus in the world: Translation English Corpus (abbr. TEC), which collects the English texts translated from other foreign language countries by British and American translators and applies the corpus to translation studies. On the basis of it, Mona Baker established the fundamental theory of translation corpus and set up the Language Engineering Research Centre for translation studies. From then on, the corpus translation studies have been prosperous.

### 1. THE THEORETICAL SOURCES OF CORPUS TRANSLATION STUDIES

The birth of Corpus Translation Studies is obviously affected by two related study fields: corpus linguistics & descriptive translation studies. The former mainly affected it from research methods and technical means, while the latter mainly affected it from theoretical frameworks and research paradigms. "Setting the corpus as the starting point of language description or the methods to verify hypothesis about language based on corpus is corpus linguistics." (Crystal, 1991, p.86) As a method for the study of language, representative sampling, corpus tagging, machine readability and the openness of corpus scale are all the important features of corpus studies. Compared with the previous small-scale, hand-manipulated and out-of-text linguistics studies, it has a lot of differences, since it collects linguistics theories, mathematical thinking mode, computer technologies and combines them together, trying to use objective and scientific methods to fully and completely describe languages under the interactions among people, computer and corpus data. After the 1980s, the optical-electronic character recognition technology greatly accelerated the process of corpus tagging treatment and promoted the analysis and utilization of corpus. Thus, corpus linguistics developed rapidly. Represented by

Birmingham English corpus, lots of corpora set up in succession, including the research on probabilistic grammar from Svartvik, J., Halliday, M. and so on. The discourse style research about British English and American English from Tottie, G. and the quantitative studies of English collocation from Sinclair J. and other scholars. Corpus linguistics shares some experience, technologies and sources with computational linguistics, but it is a completely independent subject. They are different from each other in philosophical foundation and working goals. Computational linguistics derived from Chomsky's early theory, which tries to deal with language by transformational generative grammar mechanism and it has been using the deeds all the time such as the tree structure library derived from Chomsky's early theory. But corpus linguistics comes from the empiricist traditions of field work and data collection in linguistics studies. With the help of several analysis tools such as lexical concordance, context keyword and wordlist technologies, translation studies based on corpus investigate concretely on frequency of the translation text, token, type percentage, collocation, context words, style of the translator, heterozygosis phenomena of the version, the tertiary-code phenomena and etc. what's more, the studies rigorously obey the demonstration research route of hypothesis, verification and explanation, trying to improve the scientificity of translation research by using scientific methods and form the large-scale, systemic and high-comparability research features in the end. In addition to this, the layout idea of the translation studies corpus is obviously and directly affected by classical description methods. Baker (1993) pointed out that the theoretical source of Corpus Translation Studies is not only related to the study of context and the rise of language function view in the field of linguistics, but also mainly benefits from the polysystem theory and normative theory of descriptive translation studies. Descriptive Translation Studies has unique perspectives on translation. First, translation is the fact of target culture. Sometimes, it is the fact of a certain particular identity; sometimes it may form the identifiable system belonging to itself; but all the time it belongs to target culture (Toury, 1980, p.29). Second, it is the independent text rather than the expression for other texts. Influenced by descriptive translation studies, corpus translation studies have obvious target language oriented features and the choices of text samples are strictly based on outer, temporary target language system for the most part. What's more, most of the corpus translation studies regard the translation language as an independent linguistic phenomenon to research. Certainly, it does not mean corpus translation studies will use the research mode of culturology. Because of the reason regarding the translation as a variant phenomenon to research, explanations about Descriptive Translation Studies are more about cognition rather than society and culture.

The appearance of translation studies based on corpus from contemporary linguistics perspectives formed different features between the contemporary translation studies and the previous translation studies. For example, generally speaking, the translation studies before corpus use the fragments of text as demonstrations rather than use the method of quantitative analysis. But after using corpus approaches, people can dispose a large number of texts and exactly describe discourse features of the chosen texts and translation style of the translators. Thus, they can correctly analyze the language feature differences between the texts of different styles. The quantitative analysis of corpus proves that the appearing frequencies of difference language forms are obviously different in different versions (original language text, target language text, translation version). It makes us more directly and clearly understand the features of translation language, know about the typical language usage of translation and the degree of language variation between different literal styles, and also explain the translator's specific language using habits from some aspects and special syntactic structures, etc. What's more, the previous studies are used to looking into the translation's standards starting from language habits, while a portion of corpus studies have begun to consider analyzing the effect translators' society and culture have on translation text such as the relative position of language and the role a certain literal style plays in literature polysystem by comparing the norms between original language and the target language. Thus we can say, the introduction of corpus combines the quantitative research and qualitative research, language research and culture research better in translation study fields and the research conclusions have relatively strong observability and repeatability. And this will promote the scientific research of translation's nature and features on the basis of large-scale data analysis.

---

## 2. HYPOTHESIS OF TRANSLATION LANGUAGE VARIANTS

---

So far, the translation studies based on corpus mainly covered all kinds of phenomena from the process to the production generation of translation. The main achievements involves in a lot of aspects, such as translation language features, translation shifts and specifications, translator's style, even translation and globalization. For example, House's research about German- English corpus focused on the matter how contemporary German was influenced by translation to produce the inner language change. Among them the most eye-catching one is study of the universals of translation. Translation Universals is defined to translate texts rather than the typical language features appearing in metadiscours, and the features are not the result of specific language system interference (Baker, 1993,

p.243). It means that translation language is regarded as a kind of object language variant, which entirely has some regular characteristics which are different from the source language and the rules of translation language. This definition indicates that Translation Universals as a kind of probabilistic distribution can be gained by induction and it has no relationship with the difference of language system. It is the inner feature of translation. Thus, the scholars firmly believe the studies based on this hypothesis may be a good way to reveal the nature of translation.

Before Corpus Translation Studies sprung up, lots of translation theorists have made a large amount of discussions about Translation Universals, such as Blum, Vinson, Vanderauwera, Toury, etc. They concluded three main features: simplification, which means that translators should simplify the language and information of the original work; explicitation, which means translators should clearly clarify the implicit information of the original work; and normalization, which means translators should try to make translations in line with the target language specifications to reduce text features of the source text. There are lots of achievements but obvious limitations in these studies from an overall perspective: lacking the precise identification and evaluation theory frame for the collected information; most of the studies are limited to the analysis of the sentence level and they are also mainly based on the researches of texts' literal style; the researchers collect little number of texts and use strategies which conflict each other, thus sometimes, a part of the gained conclusions are not right.

On the basis of predecessors, scholars of Corpus Translation Studies use the previous achievements, take the translation product and target-language-oriented perspective and carry on the further research combined quantitative and qualitative methods to translation universals matters. Because of the obvious characteristics of data driven type of corpus studies, its theoretical conclusions are deduced from concrete data and can be verified again and again, which effectively cover the shortage of the previous pure qualitative analysis. Since Baker (1993) put forward Translation Universals studies based on corpus, lots of studies based on the topic have sprung up. Meta specially launched a journal for Corpus Translation Studies in 1988. There are lots of representative achievements including Baker (1993, 1995, 2006), Laviosa (1997, 1998a, 1998b, 2002), Olohan, Baker, & Olohan (2003, 2004), Mauranen (2002), Mauranen, Kujamki, & Kenny (1998, 2001, 2005), Condit (2002, 2004), etc. For more than a decade, scholars have done empirical tests for Translation Universals features, in which there are demonstrations and queries. And they have made new discoveries besides the previous three features of Translation Universals: first, leveling out--- translation texts are inclined to be similar to each other in language features; second, centralization- translation texts show more convergence on Translation Universals

features (Laviosa, 1997). But until now, Translation Universals is still a hypothesis to investigate rather than an existing truth to demonstrate.

### 3. ANALYSIS MODE BASED ON THE INTERACTIONAL SOCIOLINGUISTICS

The previous linguistics approach researchers bring society into language (Nida, 1964). They believe that language contains social elements and the language of translation reflects features of different social culture. But language is not equal to society. The tasks of translation studies are trying to know the usage laws of translation language by language analysis. In the early 1990s, Mason's research also paid attention to the influence social and historical context have on text besides the effects of the texts' inner language structures. In the new century, because of the worldwide conflicts, a large number of people who are interested in translation are involved in intelligence gathering, cultural diversity coordination and propaganda activities, etc. A great amount of the translation practice was affected by ideology and participated in the activities which are against unequal rights and influence society in more obvious and ethical ways. During translation and other processes of intercultural communication, we can find lots of different narrative methods for the same event, causing people's attention of its appraisal and evaluation system hidden behind. Limited by reality, history, time, power and other elements, narrators shows some behaviors' intentions and mode at the same time of narration, which means the narration itself contains the explanation and comprehension hint of its action and offers some possibilities of retrospective study to recognize phenomena; also it promote scholars to make more new thinkings which are about translation and coordinate about language and culture boundary.

Baker's study (2006b) from Translation and Struggle about translation and social culture has changed greatly compared with her study in 1992. She holds that translation means a lot in the birth, maintenance and change of social relations. From the perspective of critical theory, she thinks the goals of translation study are to help people set up and improve the realization of society, help the translators know the importance of their work, and by revelation of ideology that the version's contextual language structure contains, readers can realize the political reality behind translation and will not be affected easily by the text's ideology. Baker mainly makes discussions about the political and constructive features of narrative concepts and connects them with the functions of translation. She cited the socialists' view and holds that every time the reception of a certain narration involves the refusal of other narrations. Realizing this, we are able to know narrations are radically political and different among people. In terms of the critical meaning,

it is the differences of the competitive narrations that give themselves meanings (Baker, 2006b, p.20; Bennett & Edelman, 1985, p.160). At the same time, she also thinks that narration is an initiative way to construct reality. The political interference of narration, different methods of narration and reception methods affect the state between defiance and control. First, the narrative definition of the past determines the contemporary narration. To compete and challenge the contemporary narration, individuals and groups choose from the narration before to highlight the obvious features of the contemporary narration. Then, the restatement of the past narration is also a way to control. It socializes individuals, makes them enter into the constructed society and political order and encourages them to use the recognized narration to explain the current events. If these are regarded as legal, it will limit the scope of the current individuals' narration and the consciousness of themselves (to choose social roles for themselves). Like Mexican-American and African-American immigrants, they all emphasize a kind of being unswervingly and excessively oppressed group consciousness through the similar restatement of the past collective narrations to dispute for their oppressions nowadays.

Because of the reason that a large number of today's conflicts are not limited to a single language group, in most of the situations, they take place in pancratium places among nations. Even some regional and local inner conflicts have to turn to typical cross-cultural and cross-linguistic ones. When a version of narration turn into another language by means of restating and translating, it will always be injected with factors outside and factors of some other more vast narration or individuals' narration factors of new environment. Thus, translation plays a very important role in the course of the narration socialization.

---

## CONCLUSION

In fact, it has been verified by not a few concrete corpus studies that the part of the automated is really very little among corpus studies of prominent characters in quantitative analysis, but more about artificial design ideas, the research foothold and qualitative analysis aiming at the nature of automation result.

Corpus studies will develop only by increasing the rational theoretical explanation and establishment on the base of description. Olohan (2004) has introduced the main problem concretely in the course of the corpus design, and points that it's the purpose of researching the problems and the hypothesis waiting for being verified that determine the standard of the corpus construction. Time and region play an extremely important role in the corpus design.

At the same time, the nationality, race, age and sex of the author and the speaker can all be standards to enter corpus, which are determined by the degree of correlation with the issues that the scholars want to investigate from corpus; and when evaluating corpus study methods,

Kenny points that the first thing to be considered is the vast noise (noise means the unrelated information) when the automated technology works, which has to be eliminated by human analysts who are interested in the lexical creation by hand. But a much harder thing is that the information people get from a fixed type corpus is far more different from what they are able to get rationally. For example, the keyword and the word string, which gained from the wordlist software's disposal of texts, vary according to the parameter values designed by users. Evaluation retrospectively studies the behaviors of abstract concepts like lexical creative usage by semi-automated methods. Finally, there is one only method, which is to compare the software result with the solution hints mechanically gained by the trained human professors. However, like Munday (2002), Mason and Serban (2003), they gain some relatively valuable research achievements with only partial help of corpus in their studies, and the key point is the concrete analysis and explanation of the observed result and the artificial design.

In a word, translation studies based on corpus are more about the quantitative analysis of the facts and are mainly about descriptions, but less about the explanation of the inner agent and the mechanism behind it.

To entirely describe the general picture of translation phenomena, it's necessary to combine the corpus studies technical methods and the other analysis tools together, to connect the discoveries with the generation process of the text and to study the quantitative research from social, cultural, historical, political and cognitive angles. This will be the further development direction of Corpus Translation Studies.

---

## REFERENCES

- Barker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In Baker et al. (Eds.), *Text and technology: In honour of John Sinclair*. Amsterdam & Philadelphia: John Benjamins.
- Bennett, W. L., & Edelman, M. (1985). Toward a New Political Narrative. *Journal of Communication*, 35(4), 156-171.
- Crystal, D. (1991). *A dictionary of linguistics and phonetics*. Oxford: Basil Blackwell Ltd.
- Halliday, M. A. K. (1991). Corpus studies and probabilistic grammar. In K. Aijmer & B. Altenberg (Eds), *English corpus linguistics: Studies in honour of Jan Svartvik*. London and New York: Longman.
- Laviosa, S. (1997). How comparable can "comparable corpora" be? *Target*, 9(2), 289-319.
- Olohan, M. (2004). *Introducing corpora in translation studies*. New York: Routledge.
- Sinclair, J. (1985). Basic computer processing of long texts. In C. Cardlin & G. Leech (Eds). *Computers and the English Language* (pp.185-203). London: Longman.
- Toury, G. (1980). *In search of a theory of translation*. Tel Aviv: Porter Institute.