



Examination on translation universals and evaluation through corpora

Xu Yuemeng

Institute of Modern History, Chinese Academy of Social Sciences, China.

Email: yfihebe@foxmail.com

Abstract

This article explores the profound impact of corpora on Translation Studies, particularly in uncovering translation universals and enhancing evaluation methods. Corpora, which are collections of authentic language data, serve as an objective foundation for translational research. The article delves into four key universal characteristics of translated texts: simplification, explicitation, normalization, and leveling out. Simplification refers to the tendency of translated texts to use less complex language, explicitation involves spelling out implicit information, normalization emphasizes conformity to target language patterns, and leveling out suggests a convergence towards typical target language features. Through corpus-based approaches, the objectivity of translation evaluation is significantly improved, as they provide concrete criteria for assessing translation quality, thereby addressing the subjectivity inherent in the field. The article emphasizes the enrichment that corpora have brought to Descriptive Translation Studies and highlights their potential to foster more objective evaluation and teaching methods. Looking towards the future, the integration of corpora in translation studies promises continued interdisciplinary advancements, propelling innovation in translation research and education.

Keywords: Corpora, Descriptive translation studies, Empirical studies, Translation evaluation, Translation universals.

1. Introduction

Translation is a broad concept that can be interpreted in various ways. It may be viewed as either a product or a process, encompassing sub-categories such as literary, technical, and machine translation. Roman Jakobson provides a crucial distinction between three types of written translation: intralingual, interlingual, and intersemiotic translation (Jakobson, 1959). Only interlingual translation, where the signs of one language (source text) are substituted with signs from another language (target text), is considered "proper translation" by Jakobson. When the message or meaning of the original text is translated into another language, the intended reception should be equivalent to that of the original audience. Consequently, translation is an activity that occurs at the language level, and linguistics is essential to the study of translation.

Until the early 1960s, linguistics was primarily characterized by descriptive research, where individual grammars were detailed but not compared, limiting their theoretical value for translators. The simultaneous emergence of two grammar theories significantly impacted the field of translation studies. The culmination of these evolving theories can be seen in Noam Chomsky's *Syntactic Structures* (1957); Eugene Nida's *Toward a Science of Translating* (1964), and Chomsky's *Aspects of the Theory of Syntax* (1965). Although Chomsky's generative transformational grammar was utilized in the development of translation studies, it was not intended as a theory of translation, and Chomsky cautioned against its misuse in that context. In addition to Chomsky's theory, numerous other linguistic theories, including those from the Prague and London schools, have had a significant influence on translation studies.

In recent years, corpus linguistics, a linguistic branch that emerged in the 1960s, has ushered in a new era in translation studies. Corpus linguistics can be defined as "the study of language based on examples of real-life language use" (McEnery and Wilson, 2001). Its primary objective is to describe language as it is actually used by speakers and to subsequently derive rules based on this usage. All observations in corpus linguistics are grounded in corpus data, and any hypotheses or generalizations about the language being studied are directly derived from this data.

A corpus is defined as "a collection of language pieces selected and ordered based on explicit linguistic criteria, serving as a sample of the language" (Sinclair, 1991). Essentially, a corpus consists of texts that are considered representative of a particular language. By definition, corpora are repositories of language as it is used in real-life situations, comprising authentic data that is highly valuable in language studies. The empirical data provided by corpora allows linguists to make objective statements based on the actual use of language, rather than subjective statements based on individual cognitive perceptions. Consequently, the unique characteristics of corpora have led to the popularity of corpus-based research, and Translation Studies is one of many disciplines that has benefited from the empirical data provided by corpora.

The use of corpora in translation studies is a relatively recent development, initiated by Mona Baker in 1993, who highlighted the significance of applying corpus evidence to examine the nature of translated texts (Baker,

1993). One compelling reason to utilize corpora in translation studies is that translation, as a unique communicative event, must be documented and analyzed as it truly occurs. Corpora offer translation theorists a distinct opportunity to observe their subject of study and identify characteristics of translated texts that can enhance their understanding of translation and its mechanisms. In fact, corpus-based approaches to translation focus on examining “how translations deviate from their source texts or from original texts written in the target language, and how particular languages, genres, and translators’ stylistic preferences impact translations” (Zanettin, Bernardini, and Stewart, 2003).

Since 1993, numerous research studies within the field of Translation Studies have employed corpus methodology to address various inquiries, ranging from the nature of translation language to the applications of different corpora types in translation education. As Laviosa (1998) notes, in recent years, “an increasing number of scholars in translation studies have started to seriously contemplate the corpus-based approach as a viable and productive perspective for studying translation and translating in a novel and systematic manner.” Consequently, the corpus-based approach has emerged as a new research paradigm (Wang, 2006).

The emergence of corpus-based translation studies has benefited several areas, including translation universals, translation evaluation, computer-assisted translation, and translator training, among others. This paper will specifically concentrate on translation universals and translation evaluation.

2. Translation Universals

Historically, translation scholars have consistently pursued generalization about the language and process of translation. The early prescriptive statements outlining what translations should be represent the initial efforts by scholars to achieve generalization. These statements all “reflect a universally valid translation ideal” (Chesterman, 2004, p. 34). However, these early attempts at generalization suffered from a significant flaw: overgeneralization. In reality, translations are not uniform, and certain prescriptive principles may only be applicable to specific types of translation. The transition from prescriptive to descriptive approaches in translation studies was facilitated by key concepts such as Frawley’s (1984) notion of translation as the third code, Blum-Kulka’s (1986) explicitation hypothesis, and Toury’s (1995) laws of translation. These concepts, characterized by their target-oriented descriptive approaches, laid the groundwork for the emergence of corpus translation studies within the field. Corpus translation studies, which grew out of corpus linguistics and is inherently aligned with linguistic approaches to translation, signifies a shift from prescriptive to descriptive approaches in translation studies.

Building on these notions, Baker initially explored the use of corpus evidence in studying translation universals in 1993. She argued that corpora can offer researchers empirical data to investigate translation universals, which she defined as “features that typically appear in translated texts rather than original utterances and are not the result of interference from specific linguistic systems” (Baker, 1993, p. 243). Baker identified four key features of translated texts, also known as translation universals: simplification, explicitation, normalization, and levelling out. The subsequent sections will concentrate on these four features and significant corpus-based research conducted on this topic.

2.1. Simplification

Simplification refers to “the tendency to simplify the language used in translation” (Baker, 1996, p. 181). This feature suggests that translated texts exhibit shorter average sentence length, reduced lexical density, and a lower lexical variety or type-token ratio compared to original texts. This occurs because employing more grammatical and fewer lexical words facilitates easier processing of the text. Additionally, when translated texts utilize a more limited vocabulary range than their original counterparts in the same language, this may be indicative of a subconscious simplification strategy employed by translators.

Numerous studies have examined the linguistic characteristics of English translated texts, with Laviosa’s (1998) corpus-based research being one of the most comprehensive. Laviosa concentrated on three potential indicators of lexical simplification: reduced lexical variety, decreased information load, and shorter sentences. To explore the linguistic nature of translated English, she utilized a subsection of the English Comparable Corpus, comprising newspaper articles. Her study revealed that translated articles exhibited a comparatively lower proportion of lexical words, a higher proportion of frequent words, and a lower average sentence length. In the same year, Laviosa investigated whether translated narrative prose shared the linguistic patterns observed in translated newspaper articles. This research also demonstrated that translations had a lower type-token ratio, reduced lexical density, and a proportional overrepresentation of the most frequent words.

While more detailed analyses and the consideration of additional factors are necessary for the next stage of universals research concerning syntax and connectors, simplification may be employed as a strategy by translators. This is because less complex sentence structures and the use of more frequently occurring words facilitate easier processing of translated text for speakers of the target language.

2.2. Explicitation

This is one of the most extensively studied and debated potential universals in translation. The concept of “explicitation” originated from Blum-Kulka’s hypothesis and was further developed by Baker as a universal feature of translation. According to Baker, explicitation denotes “a general tendency to spell out information rather than leave it implicit in translation” (Baker, 1996, p. 180). This means that the translation process often involves the addition of information and linguistic material, resulting in translations that are typically longer than their original counterparts, regardless of the languages involved. This feature can be observed at the textual, syntactical, and lexical levels.

At the textual level, text length serves as an indicator of explicitation. A notable example is provided by Johansson (as cited in Baker, 1996, p. 181), who discovered an average increase of approximately 10% in the number of words in English translations compared to Norwegian originals, as well as a slight average increase in the opposite direction (Norwegian translations compared to English originals).

At the lexical level, the tendency to make things explicit in translation may manifest through the use or overuse of explanatory vocabulary and conjunctions. This is particularly evident in Chinese-English translation, where Chinese phrases and idioms are often more culturally specific, necessitating the representation of implicit cultural information through explanatory brackets or annotations. The following examples serve to illustrate this tendency.

[2: 1] ST: 情绘一像，一手挽红丝，一手携杖悬姻缘簿，奔驰于非烟非雾中……

[2: 1a] TT: It was a picture of the old man holding, in one hand, a red silk thread (for the purpose of binding together the hearts of all couple) and, in the other, a walking-stick with the Book of Matrimony suspended from it. 林语堂译《浮生六记》)

As illustrated in examples [2: 1] and [2: 1a], the bolded words within the brackets provide explanations regarding the meaning and purpose of the "red silk thread." For Chinese readers, it is straightforward to comprehend that this sentence portrays Yue Lao, akin to Cupid in Western culture, and the "red silk thread" symbolizes the binding of hearts between a destined couple. However, readers lacking this cultural context may find it challenging to understand. Consequently, during the translation process, Lin Yutang incorporated explanatory elements to elucidate the significance of the "red silk thread" and enhance comprehension.

[2: 2] 靠山吃山，靠水吃水，我老汉靠沙子，当然要吃沙！

[2: 2a] *If* you live on a mountain, you live off the mountain; *if* you live by the water, you live off the water. I live on the sand, *and* I'll live off the sand. (王克非, 《基于语料库的英汉语比较与翻译考察》)

In Chinese, logic is often implicit and mutually understood among Chinese speakers, audiences, writers, and readers. However, when translating into English, which emphasizes structure and logic, this implicit logic must be made explicit, as demonstrated in examples [2: 2] and [2: 2a].

Similarly, explicitation also occurs in English-Chinese translation. English is rich in rhetorical devices such as repetition, alliteration, and metaphor. Some repetitions lack subjects, and metaphors may be unfamiliar to Chinese readers. Furthermore, certain cultural images may not be well-known to Chinese readers. Therefore, to ensure that translations of these instances adhere to the patterns of the target language and enable readers from different cultures to comprehend the embedded cultural messages, translators must provide necessary explanations.

[2: 3] Those were the words that were to make the world blossom for me, "like Aaron's rod, with flowers". (Helen Adams Keller, *The Story of My Life*)

[2: 3a] 后来就是这些词把一个美好的世界展现在我的面前，就像《圣经》上说的，“亚伦的杖开了花”一样。

The examples provided above serve as illustrations of explicitation. However, further study of this concept is necessary due to the numerous factors that influence explicitation, such as the language itself, social culture, the translators, and the text being translated. The role of translators extends beyond merely transferring meanings from one text to another; they must also furnish readers with background knowledge to enhance readability and comprehension.

2.3. Normalization

Normalization is defined as "a tendency to exaggerate features of the target language and conform to its typical patterns" (Baker, 1996, p. 182). This phenomenon is most apparent in the use of typical grammatical structures, punctuation, collocational patterns, and clichés. The degree of normalization is influenced by the status of the source text and language; specifically, the higher the status, the less likely normalization will occur, and vice versa. This translation characteristic is supported by research conducted by scholars such as Shlesinger (1991, as cited in Baker 1996); Ben-Shahar (1994, as cited in Baker 1996) and Kenny (2001). In her study of simultaneous and consecutive interpreting in multilingual trials, Shlesinger (1991, as cited in Baker 1996) identified numerous examples of normalization, including rounding off unfinished sentences, correcting ungrammatical utterances, and eliminating hesitations and false starts. Ben-Shahar (1994, as cited in Baker 1996) also found another instance of normalization, where "marked" structures are frequently normalized in translation.

Kenny's (2001) study is perhaps the most significant corpus-based investigation of normalization conducted to date. Utilizing a parallel corpus of contemporary German literary texts and their English translations, Kenny examines lexical creativity and normalization. The primary research question posed by the study is whether translators typically rely on more conventional target language resources to replace unconventional or text-specific lexical features in source language texts (Kenny, 2001, p. 111). In her analysis, words and collocations that occurred only once or were used by a single author were deemed potentially creative (Kenny, 2001, p. 129). The findings of Kenny's study indicate that 44% of creative word forms occurring only once in the corpus and 16% of creative collocations were normalized. Consequently, normalization can be considered a characteristic of translation. For translators, this implies a responsibility to ensure that their translations of source texts adhere to the typical features of the target language.

2.4. Leveling Out

Leveling out, as defined by Baker (1996, p. 184) refers to "steering a middle course between two extremes, converging towards the center, with the notion of center and periphery being defined within the translation corpus." This concept was derived from Shlesinger's (1989, as cited in Baker 1996) idea of the equalizing effect. However, this feature has received relatively little attention in studies of translation universals.

One of the few studies on leveling out was conducted by Shlesinger (1989, as cited in Baker 1996) which examined the effects of simultaneous interpreting on the level of orality and literacy in a text. Shlesinger discovered that during simultaneous interpreting, oral texts became more literate and literate texts became more oral, suggesting that the process of translation tends to shift texts towards the center of the continuum and away from the extremes.

3. Translation Evaluation

Translation evaluation is an inescapable issue in Translation Studies, yet assessing the quality of translations remains a contentious topic. Despite extensive debates, there is little consensus on how translations should be evaluated. This lack of agreement can be attributed to multiple factors, including the elusive and relative nature of quality, which is often influenced by social and cultural values and priorities, as well as the diverse perspectives on translation. However, the primary challenge in translation evaluation stems from its subjective nature, as the concept of quality has vague and shifting boundaries (Bowker, 2000, p. 183). Consequently, translation evaluators require objective and concrete criteria to provide compelling justifications for their assessments of good or bad translations.

The most significant contribution of corpora to translation evaluation lies in their ability to address the issue of objectivity versus subjectivity. As authentic representations of language use, corpora can offer translation evaluators concrete criteria for assessing translations. The primary advantages of corpora are their authentic data and the ease of retrieving information from them. To examine the effectiveness of corpora as a resource for translation evaluators, Bowker (2000) conducted a study comparing two groups of evaluators: one using conventional resources and the other using a corpus. The results indicated that the corpus "enabled evaluators to identify and correct a larger number of errors in a more objective manner" (Bowker, 2000, p. 206). Furthermore, Bowker (2000, p. 205) noted that "it appears that a corpus can significantly reduce the subjective element in translation evaluation, benefiting evaluators."

As demonstrated above, utilizing corpora and corpus analysis tools can assist translation evaluators in adopting a more objective approach to evaluating translations. Moreover, corpus evaluation has implications for translation teaching in the classroom. By providing objective criteria that are acceptable to both translation teachers and their students, corpus evaluation can aid teachers in delivering feedback to help students improve.

4. Conclusion

Corpora have proven to be valuable tools, and corpus linguistics techniques and methodologies have illuminated both theoretical and applied aspects of Descriptive Translation Studies. The availability of large corpora of both original and translated texts, combined with advances in corpus-based methodology, have enabled translation scholars to identify the unique features of translated language. Additionally, corpora offer translation teachers and students information that extends beyond what conventional resources provide.

The future of corpus studies in translation will be influenced by effective collaboration between corpus linguistics and computer science. For instance, recent advances in multi-modal analysis are noteworthy. It is likely that multi-modal corpus linguistics will be of great interest to scholars studying sign languages, sign language interpreting, and audiovisual translation.

Furthermore, it is reasonable to predict that the creation of web-derived mega corpora and standard-size corpora in an increasing number of languages will enhance interdisciplinary research by fostering the development of corpus-based contrastive studies, Translation Studies, and the study of language contact.

References

- Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis and E. Tognini-Bonelli (Eds.), *Text and technology: In honor of John Sinclair* (pp. 233-250). Amsterdam: John Benjamins.
- Baker, M. (1996). Corpus-based translation studies: The challenge the lie ahead. In H. Somers (Ed.), *Terminology, LSP and Translation: studies in language engineering, in honor of Juan C. Sager* (pp. 175-186). Amsterdam: John Benjamins.
- Bernardini, S. (1997). A trainee translator's perspective on corpora. *Corpus use and learning to translate*. Retrieved from <http://www.sslmit.unibo.it/introduz.htm>.
- Blum-Kulka, S. (1986). Shifts of cohesion and coherence in translation. In J. House and S. Blum-Kulka (Eds.), *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies* (pp. 17-35). Tübingen: Gunter Narr.
- Bowker, L. (2000). A corpus-based approach to evaluating student translations. *The Translator*, 6(2), 5-13.
- Chesterman, A. (2004). Beyond the particular. In A. Mauranen and P. Kujamaki (Eds.), *Translation Universals: Do they exist?* (pp. 33-51). Amsterdam/Philadelphia, PA: John Benjamins.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.
- Frawley, W. (1984). Prolegomenon to a theory of translation. In W. Frawley (Ed.), *Translation: Literary, Linguistic and Philosophical Perspectives* (pp. 159-175). London/Toronto: Associated University Press.
- Jakobson, R. (1959). On linguistic aspects of translation. In R. Brower (Ed.), *On Translation* (pp. 232-39). Cambridge, MA: Harvard University Press.
- Keller, H. (1903/2007). *The story of my life*. Whitefish: Kessinger Publishing.
- Kenny, D. (2001). *Lexis and creativity in translation: A corpus-based study*. Manchester: St. Jerome Publishing Company.
- Laviosa, S. (1998). The corpus-based approach: A new paradigm in translation studies. *Meta*, 43/44, 474-489.
- McEnery, T., & Wilson, A. (2001). *Corpus linguistics: An introduction* (2nd ed.). Edinburgh: Edinburgh University Press.
- Nida, E. A. (1964). *Toward a science of translating: With a special reference to principles and procedures involved in Bible translating*. Leiden: E.J. Brill.
- Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Stewart, D. (2000). Conventionality, creativity and translated text: The implications of electronic corpora in translation. In O. Maeve (ed.), *Intercultural Faultlines, research models in Translation Studies 1: Textual and cognitive aspects* (pp. 73-91). Manchester: St. Jerome.
- Toury, G. (1995). *Descriptive translation studies and Beyond*. Amsterdam: John Benjamins.
- Zanettin, F. (2001). Swimming in words: Corpora, language learning and translation. In A. Guy (ed.), *Learning with corpora* (pp. 177-197). Houston, Texas: Athelstan.
- 林语堂译. *Six Chapters of a Floating Life*. Shen Fu. Beijing: Foreign Language Teaching and Research Press · 1999.
- 王克非. "Corpus-based translation studies: A new research paradigm. *Foreign Languages in China* 2006, (3) : 5-6.