https://doi.org/10.16945/inahsl.24.1.5

『인문언어』(*Lingua Humanitatis*) 24권1호 (2022): 117-153

A Corpus-based Study of Translation Universals in Korean-English Literary Translation*

Woonhyung Chung
(Yonsei University)

1. Introduction

The development of the corpus-based approach in a linguistic investigation has had a significant impact on studies of translated languages. Although the translation studies prior to the advent of electronic corpora empirically found phenomena occurring in the translation process and described features of translational languages, they were mostly small-scale studies confined to specific language pairs and sentence-level analyses (Laviosa 2002:57-58). Furthermore, those studies were conducted using different texts, employing a sporadic set of questions or hypotheses; thus, their findings were not mutually comparable to support or refute the proposed hypotheses, which made it impossible to derive a unified theory (Laviosa 2002). Corpus-based translation

* This article is based on the author's master's thesis, modifying its corpus data and statistical methods and adding new indices and discussions.

studies provide solutions for such issues. Corpus linguistics allows researchers to conduct quick analyses of large amounts of texts using computer-based tools and evidence-based proposed hypotheses. More importantly, several studies have tested these hypotheses, leading to more unified definitions and a generalized theory. Therefore, translation studies have developed from a somewhat scattered and vague field to a more systematic research program, which has, in turn, produced more elaborate and sophisticated hypotheses.

Translation universals (TU) hypothesis is a prime example. Mona Baker (1993) had suggested this concept, arguing that there are "universal features" present in all translated texts. Baker indicated that as these linguistic features of translational languages are attributable to the translation process itself, the distinctive nature of translational languages is inherent and occurs independently of cultures and specific language pairs. Baker (1993, 1996) insisted that large corpora, and the development of tools and methods, provided a unique opportunity to trace the linguistic patterns of translational languages, presenting several candidates as sub-hypotheses of the TU, which included simplification, explicitation, normalization, and leveling-out. Extensive research using large-scale corpora has been conducted to investigate TU hypotheses. Laviosa (1998a, 1998b) analyzed sub-sections of the large-scale English Comparable Corpus (ECC)—newspapers sub-corpus and narrative prose sub-corpus—and verified the simplification and leveling-out hypotheses in the two text genres. Baker (2004) compared two corpora, the British National Corpus (BNC) and the Translational English Corpus (TEC), and revealed that narrative texts in translated English used more recurring and familiar lexical phrases than those in non-translated English, which provided empirical evidence to support the normalization hypothesis. Xiao (2010) contributed greatly to TU's higher generalizability by expanding the spectrum of languages in TU studies from prior studies of English translations to his study of Chinese translations. He compared

translated Chinese with non-translational native Chinese by creating the ZJU Corpus of Translational Chinese (ZCTC), and validated the simplification and explicitation hypotheses while revealing that there were some properties that were genre- or language-pair-specific, instead of following the TU hypotheses.

Although TU hypotheses have been supported by large-scale corpus-based studies, as Xiao (2010) has indicated, TU research is still confined mainly to genetically close language pairs, such as English and French or German; consequently, TU's "universality" is still difficult to claim. Thus, verification using genetically unrelated language pairs is required for TU to be a more generalized theory. For this reason, this research investigated English literary texts translated from Korean, which is not derived from a similar language tree as English. TU studies on the Korean–English pair are currently underway, but the majority of them are translations from English to Korean, and studies on translations from Korean into English are limited to a few genres. In particular, in the case of the literary genre, research is concentrated on English-into-Korean translations, whereas research on Korean-into-English translations is very scarce. Given the fact that the direction of translation in the same language pair affects translation behavior (Cheong 2006), there is a paucity of evidence from Korean-into-English translations in literary genres to support the generalizability of TU hypotheses. For these reasons, this study attempted to construct and analyze an approximately 500,000-word comparable corpus of American novels and Korean-to-English novels to verify the validity of TU hypotheses.

2. Translation Universals

The concept of TU differs from earlier approaches to translated languages in

that TU treats translations as having equal status to the source languages. As Baker (1993, 1996) noted, prior to TU hypotheses, scholars regarded translated languages as inferior to source languages—more like a second-rate activity. Even corpus linguistics excluded translated languages, even those translated by native speakers, when building their corpora. On the contrary, TU regards translation behaviors as a phenomenon that is independent of the relationship between the source and translated languages. Baker (1993, 1996) contended that there are "inherent" and universal features of translational languages; thus, the translational language should be deemed as a language with properties distinctive from the source language and/or the target language. The four following hypotheses have been proposed as the most salient characteristics of translated languages.

2.1 Simplification Hypothesis

The simplification hypothesis refers to the tendency of translated languages to be simplified at lexical, syntactic, and/or stylistic levels (Baker 1996). Simplification postulates that translators use simpler and easier words and sentences in the translated text to help readers understand it better and reduce ambiguity in interpretations. Blum-Kulka and Levenston (1983) provided initial empirical evidence in Hebrew-English translations, which supported simplification at the lexical level. They suggested that translators were more inclined to replace words in the source language with hypernyms, more familiar words, or words with little cultural connotation. Vanderauwera (1985) focused on syntactic and stylistic simplifications in translations. In her study on the English translations of 50 Dutch novels, she revealed the tendency of translators to render complex and sophisticated sentences more accessible and

readable by breaking them down into shorter and simpler forms of expressions or sentences. Malmkjær (1997) noted the use of punctuation in translations. Stronger punctuations, such as periods replacing commas or semicolons, often occurred in translations, resulting in translations having shorter and less complex sentences that were easier to read. As the beginning of a large-scaled corpus-based TU study, Laviosa (1998a, 1998b) explored a multi-source-language comparable corpus of English: the English Comparable Corpus (ECC). She analyzed its subsection on newspaper articles and narrative prose and verified the lexical simplification hypothesis of translational languages having lower lexical density (LD), a higher proportion of high-frequency words, and a fewer number of lemmas in high-frequency words. Moreover, the results from newspaper articles showed that the mean sentence length (MSL) of translations was shorter than non-translations, which supported the syntactic simplification hypothesis; however, the results from narrative prose showed the opposite trend. These contradictory results left room for the debate regarding whether sentence length could be an indicator of TU hypotheses. Since then, studies that expanded the scope of TU studies beyond English and European language pairs have been conducted to test the validity of the TU hypothesis. Xiao (2010) found that translated Chinese from English had a lower lexical density and a higher proportion of high-frequency words than native Chinese from English in 15 genres, which showed that the lexical simplification hypothesis can be applied to English-Chinese translations. Furthermore, as with the results of Laviosa's studies (1998a, 1998b), MSL, which is an aspect of the syntactic hypothesis, was sensitive to genre variation and yielded inconsistent results, indicating that it is a less reliable index. Lee (2013) analyzed the English-Korean literary corpus and supported both the lexical and syntactic simplification hypotheses by showing that the translation corpus was lower in both the standardized type-token ratio (STTR) and MSL. The inconsistent results for

MSL in previous research show the need for verification which broadens the scope of language pairs and genres in the TU study.

2.2 Explicitation Hypothesis

The explicitation hypothesis relates to the tendency in translations to "spell out things rather than leave them implicit" (Baker 1996). It postulates that translated texts would display a higher incidence of lexical, grammatical, or semantic devices than source texts, as translators have a (sub)conscious intention to express meanings and/or thoughts of characters more explicitly. The devices for the increased explicitness and transparency of translation texts can be exemplified by inserting interjections, modifiers, qualifiers, connectives, and grammatically selective words, adding detailed explanations and descriptions, and disambiguating pronouns with more precise referring expressions (Vanderauwera 1985).

The frequency of connectives has been studied as evidence of the explicitation hypothesis (Halverson 1996; Puurtinen 2004; Chen 2004, 2006; Xiao and Yue 2009; Xiao 2010). In genre-specific studies, Puurtinen (2004) found that clause connectives were used more frequently in translated Finnish children's literary texts than in non-translated ones. Chen (2006) revealed that connectives were more frequent in popular science books in English-to-Chinese translations than in their native Chinese counterparts. In his across-genre study, Xiao (2010) found that although there were a few differences between genres, conjunctions tended to occur more frequently in translated Chinese than in their native counterparts. Furthermore, Xiao (2010) revealed that translated texts favored the use of conjunctions in the high-frequency band rather than those in the low-frequency band, which he indicated manifested the explicitation

hypothesis in combination with the simplification hypothesis.

Optional syntactic elements have been used more frequently in translations as a part of expressing grammatical relations more explicitly. Burnett (1999) found that translated texts contained a larger number of optional *that* with verbs such as *suggest, admit, claim, think, believe, hop*e, and *know* than non-translated texts. Olohan and Baker (2000) also found that the optional *that* with the reporting verbs *say* or *tell* was more frequent in translated texts by analyzing the TEC and a comparable sample from the BNC. Furthermore, using the identical corpora, Olohan (2001) investigated various optional syntactic elements, such as the optional *that* with the verb *promise*, the relative pronoun *which*, the complementizer *to* following the verb *help*, *while* in the *while* + *v-ing* construction, and *after* in the *after* + *v-ing* or *after* + *v-ed* constructions, all of which were more frequent in the TEC than in the BNC.

Readability has been studied as an aspect of explicitation (Toury 1995; Øverås 1998). Toury (1995) confirmed an apparent correlation between explicitness and readability, connecting a higher level of explicitness of translated texts to the translator's (un)conscious strategy of helping readers' readability and understanding. However, Pastor et al. (2008) regarded the readability index as one of the indices for the simplification hypothesis, indicating that readability was facilitated by simple words, clauses, and sentences. The two arguments have little conflict; however, increased readability of translated texts can likely be accommodated by the two hypotheses.

On the other hand, the explicitation hypothesis predicts that translated texts have longer sentences because they tend to be informative and sometimes redundant for enhancing readers' understanding. In contrast, the simplification hypothesis expects that translated texts would have shorter and simpler sentences rather than complex and sophisticated ones. Interesting part of this is

that these two conflicting hypotheses assume that shortening or lengthening sentences help readers understand translated texts better.

2.3 Normalization Hypothesis

The normalization hypothesis, also referred to as "conventionalization," relates to the "tendency to conform to patterns and practices that are typical of the target language, even to the point of exaggeration" (Baker 1996). It posits that translators are likely to "opt for safe, typical patterns of the target language and shy away from creative or playful uses" (Baker 2007); thus, translated texts would appear to be plainer, more typical, and more familiar, having fewer peculiar punctuations, ungrammatical syntax, and idiosyncratic expressions than non-translated texts.

One of the most common indices employed to uncover the effect of normalization is the use of prefabricated and recurring phrases (Baker 2004; Kenny 1998, 1999, 2000, 2001; Lee 2011a, 2011b; Nevalainen 2005; Olohan 2004; Øverås 1998; Xiao 2011). Kenny (1999) examined *hapax legomenon*, which is a sequence of words that occurs once in the entire text, and found that it appeared less frequently in translated texts than in non-translated texts, indicating translators' avoidance of creative or idiosyncratic linguistic forms. Baker (2004) investigated fixed or semi-fixed lexical bundles by analyzing the TEC, which is a sub-corpus ECC. The finding was that fixed or semi-fixed lexical bundles, such as *that is to say*, and *in other words*, were used far more frequently in translated texts than in non-translated texts, which she indicated was the translators' intentional choice to create the effect of fluency.

Common findings regarding punctuation suggest that translated texts tend to use more normal punctuations instead of unusual peculiar ones, and display

stronger punctuations (e.g., periods instead of semicolons, and periods or semicolons replacing commas) (Vanderauwera 1985; Malmkjær 1997; May 1997). This adjustment gives translated texts clear marking breaks and higher readability compared to source texts which are "fragmented, incomplete, vague and ambiguous" (Scott 1998). In a similar vein, as a result of the translators' tendency to make safe choices, unusual or ungrammatical syntactic structures and expressions in source texts, which may have the effect of tension and suspense, tend to be replaced with plainer, more familiar, and grammatically correct equivalences (Vanderauwera 1985; Leuven-Zwart 1989, 1990; Shlesinger 1991; May 1997). The use of stronger punctuation in translations is used as evidences for two different hypotheses. Normalization hypothesis attributes it to translators' preference for clear marking breaks instead of incomplete and vague ones; however, simplication hypothesis explains that it is the result of simplifying the translated texts with shorter and less complex clauses and sentences.

2.4 Leveling-out Hypothesis

The leveling-out hypothesis postulates that translated texts tend to be more homogeneous than non-translated texts (Baker 1996). This hypothesis assumes that the translation process necessarily reduces the individual characteristics of source texts; consequently, translated texts possess lesser diversity and greater homogeneity than non-translated texts. While the normalization hypothesis posits that translated texts conform to the norm of a target language, the leveling-out hypothesis refers to the tendency of translated texts to "gravitate toward the center of a continuum rather than the fringes" (Baker 1996).

This hypothesis started to gain attention after the advent of corpora. Corpus

linguistics has made it possible to represent a level of variation as a numerical value, which enabled easy verification of the hypothesis. Verification of the leveling-out hypothesis has been conducted mainly with standard deviations of the simplification indices (e.g., LD, STTR, MSL, etc.) (Laviosa 1998a, 1998b; Lee 2013; Grabowski 2015). Laviosa (1998a) compared the standard deviations of type/token ratio (TTR), LD, and MSL, and confirmed the leveling-out hypothesis in a study on newspaper articles, yet her subsequent research (1998b) on literary genres invalidated the hypothesis in the analyses of standard variations of LD, the proportion of high-frequency words, and the MSL. However, Lee (2013) and Grabowski (2015) used the values of simplification indices (standard variations of STTR and MSL) and validated the leveling-out hypothesis in literary genres.

3. Method

3.1 Corpus

The current study constructed a monolingual comparable corpus to compare translated English from Korean with non-translated English in the literary genre. This corpus, which was named the Comparable Corpus of American and Korean Novels (CCAKN), comprises two sub-corpora: a collection of 25 American novels (AN) and a collection of 25 Korean novels translated into English (KN) (see Appendix 1). The composition of each sub-corpus is illustrated in Table 1.

Table 1. Composition of CCAKN

Sub-corpora	Number of novels	Size (tokens)	Range of publication date (year)	Range of released date on the website (AN)/ publication date of translations (KN) (year)
AN	25	250,493	1881–1925	1992–2021
KN	25	250,304	1918–2009	2002–2010

For a total of approximately 500,000 tokens, 10,000 words per novel were extracted. The fragments were randomly extracted from the entire text, excluding titles and/or title numbers. To minimize the impact of individual writers' or translators' characteristics on the entire corpus analysis results, it was limited to no more than two works per writer or translator. Regarding time span, American novels were published between 1881 and 1925, and Korean novels, between 1918 and 2009. The difference in the period between the two sub-corpora was caused by practical constraints on obtaining electronic texts. While this study obtained electronic documents for the Korean novels that had already been manually built, the electronic documents for American novels were extracted from the Gutenberg project webpage (http://www.gutenberg.org), where contains digitalized versions of books whose copyright has expired; thus, the AN is comprised of relatively older novels as compared to the KN. Although the range of publication dates has a discrepancy between the two sub-corpora, this research decided to accept this limitation considering the following three facts: (a) both the release date on the website and the publication date of translations are similar (around the year 2000), as seen in Table 1, (b) there has been little change in the American English syntax since the 19th century, and (c) time-specific words can be negligible because they will be relatively infrequent in the entire literary work.

3.2 TU Indices and Data Analysis

For comparability in verifying the four major sub-hypotheses, the study selected equivalent indices and criteria commonly employed in previous TU studies (Baker 2004; Laviosa 1998a, 1998b; Lee 2013; Xiao 2010). (see Table 2). To validate the simplification hypothesis, this study measured two types of lexical density—Stubbs-style lexical density and standardized type-token ratio (STTR)—and mean sentence length (MSL). Additionally, frequency profiles were analyzed to compare the distribution of words used in the two corpora. Regarding the validation of the explicitation hypothesis, this study examined the normalized frequencies of connectives and highly frequent connectives in the two corpora. Furthermore, following Xiao's (2010) study, the normalized frequency of connectives of six different frequency bands were investigated. To evaluate the normalization hypothesis, referring to Baker's (2004) and Mahlberg's (2007) studies, the numbers and types of three- to five-word lexical bundles were compared. Additionally, types of lexical bundles were examined. Finally, to verify the leveling-out hypothesis, the study computed the standard deviations of two types of lexical density and one of MSL.

Table 2. TU Indices selected for the sub-hypotheses

Sub-hypothesis	Indices
Simplification	Two types of lexical density (i.e., Stubbs-style lexical density and standardized type-token ratio)
	Mean sentence length
	Frequency profiles
Explicitation	Normalized frequency of connectives and high-frequency connectives
	Distribution of connectives across frequency bands
Normalization	Normalized frequency of multi-word (three to five) lexical bundles
	Types of lexical bundles
Leveling-out	Standard deviations of two types of lexical density and mean sentence length

Various quantitative measurements were used to measure the selected TU indices. First, WordSmith Tools 6.0 (Scott 2011) automatically computed STTR and MSL using its WordList function and generated the lists of connectives that occurred in each file by using its Concord function. Second, AntConc 3.5.8 (Anthony 2019) was employed to compile the lists of function words (Nation, 2001:430-431) to calculate the scores of Stubbs-style lexical density and the lists of multi-word (three to five) lexical bundles occurring in each text. Lastly, SPSS calculated the values of standard deviations of Stubbs-style lexical density, STTR, and MSL and conducted independent sample t-tests to determine statistical significance.

4. Results and Discussion

4.1 Simplification

4.1.1 Lexical Density and Mean Sentence Length

Lexical density and mean sentence length (MSL) were gauged to verify the simplification hypothesis in Korean-English literary translations. The hypothesis posits that translational languages are lexically and syntactically simpler, so the KN corpus was predicted to have a lower lexical density and a shorter MSL compared to the AN corpus.

The current study took two types of measures to gauge lexical density: Stubbs-style lexical density (Stubbs 1986:33, 1996:172) and standardized type-token ratio (STTR). Stubbs-style lexical density is the ratio of the number of word types to the total number of word tokens; STTR is a developed

method of TTR, which is calculated by averaging TTRs of consecutive segments of a text divided by a certain number of word tokens to avoid distortion from the total number of word tokens of texts (Scott 1997). While Laviosa (1998a, 1998b) took the Stubbs-style approach for her analyses, Xiao (2010) compared both measures in native and translated Chinese, pointing out that whereas Stubbs-style lexical density gauges the information load, STTR measures lexical variability. To diversify the analysis and compare it with previous research, this study compared both Stubbs-style lexical density (henceforth, LD) and STTR.

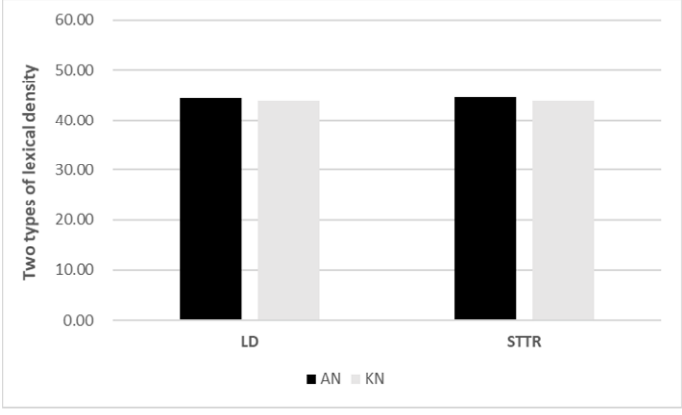

Figure 1. Stubbs-style lexical density and standardized type-token ratio in AN and KN

Figure 1 shows the LD and STTR scores in the AN and KN corpora. Notably, while the mean LD in AN (44.49%) was slightly higher than that in KN (43.88%), the mean difference (0.611) was not statistically significant ($t(36.588) = 0.663$, $p = .511$). Likewise, the AN corpus had a slightly higher STTR than the KN corpus (44.63 vs. 43.78), but the mean difference (0.845) was not statistically significant ($t(48) = 1.031$, $p = .308$).

These results contradicted lexical simplification hypothesis and results of

previous studies. Previous studies verified lexical simplification hypothesis by showing lower LD of translated English in newspaper articles and narrative prose (Laviosa 1998a, 1998b) and lower LD and STTR of translated Chinese in 15 genres (Xiao 2010). A study on English-into-Korean literary translations (Lee 2013) also supported the lexical simplification hypothesis, showing that Korean translated from English had lower STTR than non-translated Korean. Although LD and STTR have been proven to be among the most reliable indices for the TU hypothesis, the results of the current study showed no difference in both LD and STTR in the KN and AN corpora, implying that language pairs and the direction of the translation could still be important factors affecting the results of lexical simplification.

Next, the MSL scores were compared in the AN and KN corpora to test syntactic simplification hypothesis. Figure 2 shows the MSL scores in the two corpora. As can be seen, the MSL in AN (22.58) was considerably longer than that in KN (14.46). The mean difference of 8.120 was statistically significant ($t(31.212) = 6.550$, $p < .001$).

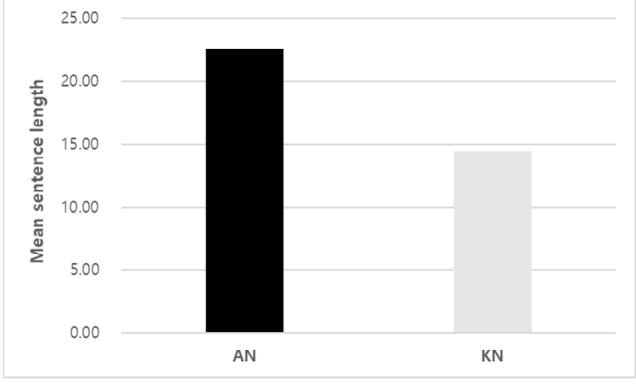

Figure 2. Mean sentence length in AN and KN

This result validated syntactic simplification hypothesis and was consistent with Lee's (2013) findings from English-into-Korean translations in the literary genre and Xiao's (2010) observation of English-into-Chinese translations in general fiction. Although Xiao (2010) revealed in analyses by genre that MSL was too sensitive to genre variation to be a reliable indication for syntactic simplification hypothesis, the result from general fiction showed that MSL was statistically significantly shorter in translated than in native Chinese fiction. In contrast, Laviosa's (1998b) study on translated English from multi-source-language in the literary genre showed the contradictory results: translated texts had a shorter MSL than non-translated texts. A possible explanation for these divergent results might be sought in the fact that translated languages simultaneously have opposite tendencies: one toward the simplification hypothesis and the other toward the explicitation one. Translations tend to use a simpler form of language (the simplification hypothesis); concurrently, they tend to contain a more informative and redundant form of language to enhance the clarification of meanings and, consequently, readers' understanding (the explicitation hypothesis).

4.1.2 Frequency Profiles

Taking a closer look at the distribution of words used in the two corpora, the present study investigated the proportions of both high- and low-frequency words. The simplification hypothesis postulates that translational texts repeatedly use more high-frequency words while using fewer low-frequency words, which tend to be creative and author-specific words. This indicates a lack of word variety in translational texts. Therefore, the KN corpus was predicted to have a higher proportion of high-frequency words and a lower

proportion of low-frequency words.

For high-frequency words, every word type that covered more than 0.10% of the total number of word tokens in a corpus was counted (Laviosa 1998b; Xiao 2010). For low-frequency words, every word type that individually accounted for less than 0.01% of the total number of word tokens in a corpus was counted. The numbers of types and lemmas in each frequency band were also computed.

Table 3. Frequency profiles of AN and KN

	AN	KN
Proportion of high-frequency words (%)	53.66	**55.41**
Number of types of high-frequency words	117	**138**
Number of lemmas of high-frequency words	95	**111**
Proportion of low-frequency words (%)	**23.09**	19.65
Number of types of low-frequency words	**16,608**	13,628
Number of lemmas of low-frequency words	**12,073**	10,036

Table 3 shows the frequency profiles of the AN and KN corpora. The high-frequency words covered a greater percentage of text in the KN corpus (55.41%) than in the AN corpus (53.66%), which provided evidence of the simplification hypothesis. The results suggested that the use of words in translated languages tended to concentrate on words in the band of high-frequency words.

Additionally, the results showed that the numbers of both types and lemmas of high-frequency words were higher in the KN corpus (138 and 111, respectively) than in the AN corpus (117 and 95, respectively). The results diverged from previous studies that observed the numbers of types or lemmas in frequency profiles. Previous studies on types or lemmas of high-frequency words have predicted and revealed that while having a higher proportion of high-frequency words, translated texts had fewer numbers of types and lemmas

of them. Laviosa (1998b) proved that translated languages had fewer lemmas in the band of high-frequency words than non-translated languages, which she insisted showed a stronger tendency toward simplification. Xiao (2010) confirmed Laviosa's assumption by investigating the number of "types" of high-frequency words because Chinese is a non-inflectional language, so that its words do not have lemma forms. The results were in line with those obtained by Laviosa (2001) in that it showed that translational Chinese had fewer word types of high-frequency words than non-translational Chinese. The present study, though, did not confirm the strong tendency toward simplification.

When it comes to low-frequency words, the result supported the simplification hypothesis. The proportion of low-frequency words was lower in the KN corpus (19.65%) than in the AN corpus (23.09%), and the numbers of both types and lemmas were also lower in the KN corpus (13,628 and 10,036, respectively) than in the AN corpus (16,608 and 12,073, respectively). This result confirmed the simplification hypothesis in that translated texts were composed of a less diverse range of words as they used fewer low-frequency words.

4.2 Explicitation

4.2.1 Proportions of Connectives and High-frequency Connectives

To test the explicitation hypothesis, this study compared the proportions of connectives in the two corpora. The explicitation hypothesis concerns that translated texts would show more frequent uses of connectives than in non-translated texts because translators (sub)consciously reveal the relations between sentences and/or paragraphs more explicitly. Based on this postulation,

the KN corpus was expected to have a greater number of connectives than the AN corpus. This study used a list of 110 connectives, which was created by consulting Quirk et al.'s (1985) and Biber et al.'s (1999) lists.

As seen in Figure 3, the mean normalized frequency of connectives was significantly higher in the translational corpus (90.59 instances per 10,000 tokens) than in the native corpus (68.90 instances per 10,000 tokens) ($t(48) = -2.724$, $p = .009$, mean difference = –21.694). This result validated the explicitation hypothesis.

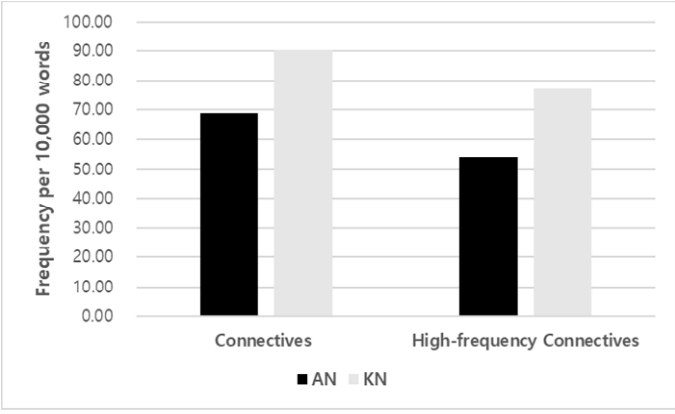

Figure 3. Normalized frequencies of connectives and high-frequency connectives in AN and KN

In addition to the overall frequency of connectives, the present study explored the highly frequent connectives in the two corpora. Previous TU studies on translated Chinese (Chen 2006; Xiao and Yue 2009; Xiao 2010) investigated connectives of different frequency bands between translated and native Chinese, and they found that connectives of high-frequency bands (i.e., connectives with a proportion greater than 0.01%) tended to cover a greater percentage of text in translated Chinese than in native Chinese. The results

from this study replicated those of previous studies, showing that the mean frequency of high-frequency connectives (i.e., connectives with a proportion greater than 0.01%) was significantly higher in the translational corpus (77.09) than in the native corpus (54.28) ($t(48) = –3.110$, $p = .003$, mean difference = –22.804) (see Figure 3).

4.2.2 Distribution of Connectives across Frequency Bands

To examine the distribution of connectives, this study explored the proportion of connectives of six frequency bands. The previous studies (Chen 2006; Xiao and Yue 2009; Xiao 2010) found that translated languages tended to display lower proportions of less frequent connectives (i.e., connectives with a proportion less than 0.01%) than non-translated languages. The results from this study were in accordance with their results (see Table 4).

Table 4. Distribution of connectives across frequency bands (proportion, %)

Frequency bands	0.05	0.01	0.005	0.001	0.0005	0.0001
AN	0.423	0.120	**0.067**	**0.067**	**0.005**	**0.007**
KN	**0.582**	**0.190**	0.032	0.060	0.003	0.006

As can be seen, connectives in the usage bands 0.05 (i.e., connectives with a proportion less than 0.05%) and 0.01 (i.e., connectives with a proportion less than 0.01%) covered a greater percentage of text in KN than in AN, but the trend was reversed for the usage band 0.005 (i.e., connectives with a proportion less than 0.005%) and below. This result showed that translated texts used a greater number of frequent connectives and a smaller number of less frequent connectives. As Xiao (2010) indicated, this tendency confirmed the findings from the frequency profile in the simplification hypothesis,

showing the distinctive tendency of favoring the use of words in a high-frequency band in translated texts.

4.3 Normalization

4.3.1 Proportions of Lexical Bundles

To verify the normalization hypothesis, this study compared the numbers and types of lexical bundles in the two corpora. The normalization hypothesis posits that translators tend to conform to a norm of a target language, and as part of this tendency of conformity, translated text uses more typical or prefabricated sequences of words instead of creative expressions. Therefore, it was predicted that the KN corpus would display a greater number of recurring lexical bundles, and their types would be more fixed forms compared to the AN corpus.

This study explored three-, four-, and five-word lexical bundles. The range of lexical bundles that should be counted seems to depend on the researchers' judgment. On the one hand, this study excluded two-word sequences based on Biber et al.'s (2002:444) indication that "two-word sequences are generally too short and numerous to be interesting." On the other hand, five-word lexical bundles were included. This is because although Biber et al. (ibid.) indicated that "five-word and six-word bundles can be found, but they are much less common," Mahlberg (2007) revealed in her corpus-based literary study that five-word lexical bundles occurred frequently enough across a broad range of texts in the literary genre. Moreover, Baker's (2004) study, which is among the most representative corpus-based TU studies, investigated the numbers of predetermined fixed lexical bundles, and its list included three- to five-word

lexical bundles. For these reasons, this study examined three- to five-word lexical bundles. In addition, contraction forms were treated as single words following Biber et al.'s (1999) study (e.g., *won't able to* was treated as a three-word lexical bundle).

In terms of cut-off, the study extracted lexical bundles that occurred at least 40 times per million (Biber and Barbieri 2007). By this standard, the frequency cut-off was set to 10 (formula: 250,000 [total number of word tokens] × 0.00004 = 10). Additionally, the range cut-off was set to 10, so lexical bundles that occurred in more than 10 novels among 25 novels were extracted.

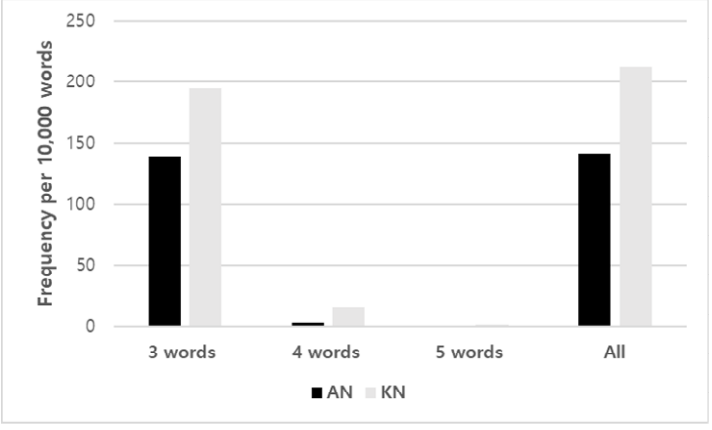

Figure 4. Normalized frequencies of lexical bundles in AN and KN

Figure 4 shows the frequencies of the three-, four-, and five-word lexical bundles and their summation frequencies. As can be seen, the mean frequencies of each bundle were higher in the translated corpus than in the native corpus (138.65 vs. 194.62 for three-word lexical bundles, 3.11 vs. 16.14 for four-word lexical bundles, and 0.00 vs. 1.75 for five-word lexical bundles, respectively). Statistical analyses confirmed these results ($t(48) = -8.491$, $p < .001$, mean

difference = –55.968; $t(34.076) = -11.387$, $p < .001$, mean difference = –13.030; $t(24) = -6.771$, $p < .001$, mean difference = −1.745, respectively). In the case of five-word lexical bundles, the bundles that satisfied the cut-off did not occur in the AN corpus, indicating that non-translated texts used fewer fixed phrases. The summation of their frequency was significantly higher in the translational corpus than in the native corpus (141.76 vs. 212.51) ($t(48) = -10.480$, $p < .001$, mean difference = –70.743). These results confirmed the normalization hypothesis.

4.3.2 Types of Lexical Bundles

When examining types of lexical bundles (see Appendix 2), the lexical bundles such as *one of the, out of the, there was no, it was a,* and *as soon as* occurred commonly in the two corpora. In examining the types that occur in one corpus only, while the types occurring only in the AN corpus tended to be related to the tense and aspect of verbs, such as *she had been* and *would have been,* rather than fixed phrases, the types occurring only in the KN corpus tended to be prepositional or adverb phrases, such as *in the middle of, for the first time, for a long time, for a while, the sound of,* and *the back of,* which were fixed or semi-fixed lexical bundles. This finding accorded with Baker's (2004) study that translated texts possessed more fixed lexical bundles than non-translated texts.

4.4 Leveling-out

4.4.1 Standard Deviations of Two Types of Lexical Density and Mean Sentence Length

The leveling-out hypothesis predicts that translational languages are more

homogeneous and have a lower level of variation. To verify this hypothesis, the study gauged standard variations of two types of lexical density—LD and STTR—and MSL in AN and KN and expected that KN would have lower standard variations of these three indices.

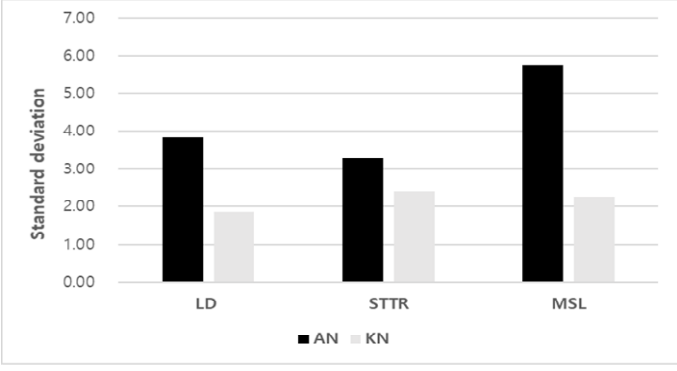

Figure 5. Standard deviations of Stubbs-style lexical density, standardized type-token ratio and mean sentence length in AN and KN

Figure 5 shows the standard deviations of LD, STTR, and MSL in the AN and KN corpora. As predicted, the non-translated AN corpus showed higher values in the standard deviations of all three indices than the translated KN corpus (LD: 3.83 vs. 1.87; STTR: 3.31 vs. 2.41; MSL: 5.77 vs. 2.26). The difference was greatest for MSL (3.51), followed by LD (1.96), and STTR (0.90).

More importantly, although the scores of LD and STTR in the AN and KN corpora were not significantly different in the verification of the lexical simplification hypothesis (see Section 4.1.1), their standard variations were lower in KN than in AN, which showed that KN was more homogeneous than AN. These results mean that apart from the simplification hypothesis, the leveling-out hypothesis is an independently valid hypothesis.

Previous research on English-into-Korean literary translations (Lee 2013) also supported the leveling-out hypothesis, showing that Korean translated from English had lower standard variations in STTR and MSL values than non-translated Korean. In contrast, Laviosa's (1998b) study on translated English from multi-source-language in the literary genre invalidated the leveling-out hypothesis as it yielded standard variations of three simplification indices (i.e., LD, proportion of high-frequency words, and MSL), which were higher in translated texts than in non-translated texts. These conflicting results may be caused by the different compositions of the translated corpora. The studies, which supported the leveling-out hypothesis (i.e., Lee's (2013) study and the current study), analyzed the mono-source-language translated corpus, whereas Laviosa's (1998b) study, which invalidated the hypothesis, was used a multi-source-language corpus. Thus, the multi-source-language of a translated corpus in Laviosa's study may have brought in the higher value of standard variation; subsequently, the variety of source languages may be responsible for the discrepancy in verifying the leveling-out hypothesis in the literary genre. The further examination is needed on whether the composition of translated corpus—mono-source-language corpus or multi-source-language corpus—affects the standard variations.

5. Conclusion

This study was conducted to test whether the translation universals can be applied to Korean-to-English literary translations. Even though the hypothesis was named as "universals," the verification of its universality has been constrained mainly to language pairs derived from similar language trees, such

as English and closely related European languages; thus, it could not be considered to have achieved its real universality. To explore the generalizability of TU, the current study attempted to investigate a genetically distant language pair, Korean-into-English translations, which have remained unexplored in TU studies. To this end, this study constructed the Comparable Corpus of American and Korean Novels (CCAKN) with a size of approximately 500,000 words and compared translated English from Korean with non-translated English in the literary genre. With the CCAKN, indices of four major sub-hypotheses—simplification, explicitation, normalization, and leveling-out—were analyzed.

To validate the simplification hypothesis, two types of lexical density—Stubbs-style lexical density and standardized type-token ratio—and mean sentence length were measured. The lexical density scores in both sub-corpora were not significantly different, which did not confirm the hypothesis. The results from the mean sentence length supported the hypothesis, showing that translated English used shorter sentences. In the examination on the composition of lexical use, the translated texts had a higher percentage of high-frequency words and a lower percentage of low-frequency words, which indicated that translated texts displayed simpler and less diverse words. In the case of the numbers of types and lemmas, even though some previous studies discovered a stronger tendency toward simplification by finding the fewer numbers of types or lemmas of high-frequency words in translated texts, the present results did not corroborated that tendency. To verify the explicitation hypothesis, this study calculated the proportions of connectives and found that they occurred more frequently in translated texts than in non-translated texts, which validated the explicitation hypothesis. More interestingly, this study found that the connectives in the high-frequency band (i.e., connectives with a proportion less than 0.01%) covered more areas of translated texts than of

non-translated texts, revealing that the simplification occurs in combination with the explicitation. The evaluation of three-, four-, and five-word lexical bundles confirmed the normalization hypothesis. Their occurrences were more frequent in the translated texts than in the non-translated texts; moreover, their types in the translated texts tended to be more fixed and typical than those in the non-translated texts. Finally, the standard deviations of two types of lexical density and mean sentence length were calculated to test the leveling-out hypothesis, with lower scores in translated texts indicating more homogeneity than in the non-translated texts.

All the results, except for lexical density, corroborated the TU hypotheses. The finding that the TU was considerably applicable to Korean-to-English translations—translations from a genetically distant language pair—increased the possibility of its universality. Although the current study has a limitation in that the corpus covered only the literary genres, the study has contributed to verifying the TU, considering that the Korean-to-English translations in the literary genre had remained unexplored. Similar to the English Comparable Corpus and the ZJU Corpus of Translational Chinese, more balanced corpora containing different genres are required for TU studies of Korean-to-English translations.

Reference

Anthony, L. *AntConc (Version 3.5.8)* [Computer Software]. Tokyo, Japan: Waseda University, 2019. Available from https://www.laurenceanthony.net/software

Baker, M. “Corpus Linguistics and Translation Studies. Implications and Applications.” *Text and Technology: In Honour of John Sinclair.* Eds. M. Baker, G. Francis & E. Tognini-Bonelli. Amsterdam: John Benjamins, 1993. 233-50.

Baker, M. “Corpus-based Translation Studies: The Challenges that Lie Ahead.” *Terminology, LSP and Translation.* Ed. H. Somers. Amsterdam: John Benjamins, 1996. 175-86.

Baker, M. “A Corpus-based View of Similarity and Difference in Translation.” *International Journal of Corpus Linguistics* 9:2. (2004). 167-93.

Baker, M. “Patterns of Idiomaticity in Translated vs. Non-translated Text.” *Belgian Journal of Linguistics* 21:1. (2007). 11-21.

Biber, D., & Barbieri, F. “Lexical Bundles in University Spoken and Written Registers.” *English for Specific Purposes* 26. (2007). 263-86.

Biber, D., Conrad, S., & Leech, G. *Longman Student Grammar of Spoken and Written English*. Harlow: Longman, 2002.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. *Longman Grammar of Spoken and Written English.* New York: Longman, 1999.

Blum-Kulka, S., & Levenston, E. “Universals of Lexical Simplification.” *Strategies in Interlanguage Communication.* Eds. C. Faerch & G. Kasper. New York: Longman, 1983. 119-39.

Burnett, S. *A Corpus-based Study of Translational English*. MSc Thesis, Manchester: Centre for Translation and Intercultural Studies UMIST, 1999.

Cheong, H. J. “Target Text Contraction in English-into-Korean Translations: A Contradiction of Presumed Translation Universals?” *Meta: journal des traducteurs/Meta: Translators' Journal* 51:2. (2006). 343-67.

Chen, W. “Investigating Explicitation of Conjunctions in Translated Chinese: A Corpus-based Study.” *Language Matters* 35:1. (2004). 295-312.

Chen, W. *Explicitation through the Use of Connectives in Translated Chinese: A Corpus-based Study.* Unpublished PhD Thesis, Manchester: Manchester University, 2006.

Grabowski, Ł. “On Translation Universals in Selected Contemporary Polish Literary Translations.” *Studies in Polish Linguistics* 7:1. (2015). 165-83.

Halverson, S. “Norwegian-English Translation and the Role of Certain Connectives.” *Translation and Meaning Part 3.* Eds. A. Thelen & B. Lewandowska-Tomaszczyk. The Netherlands: Hogeschool Maastricht, 1996. 129-39.

Kenny, D. “Corpora in Translation Studies.” *Routledge Encyclopedia of Translation Studies.* Ed. M. Baker. London: Routledge, 1998. 50-53.

Kenny, D. “The German-English Parallel Corpus of Literary Texts (GEPCOLT): A Resource for Translation Scholars.” *Teanga* 1. (1999). 25-42.

Kenny, D. “Translation at Play: Exploitations of Collocational Norms in German-English Translation.” *Working with German Corpora.* Ed. B. Bodd. Birmingham: University of Birmingham Press, 2000. 143-60.

Kenny, D. *Lexis and Creativity in Translation: A Corpus-based Study.* Manchester: St. Jerome Publishing, 2001.

Laviosa, S. “The English Comparable Corpus: A Resource and a Methodology.” U*nity in Diversity? Current Trends in Translation Studies.* Eds. L. Bowker, M. Cronin, D. Kenny & J. Pearson. Manchester: St. Jerome Publishing, 1998a. 101-112.

Laviosa, S. “Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose.” *Meta: Translators’ Journal* 43:4. (1998b). 557-70.

Laviosa, S. *Corpus-based Translation Studies: Theory, Findings, Applications*. Amsterdam: Rodopi, 2002.

Lee, C. “Corpus-based Analysis of Lexical Bundles in Literary Translation.” *Interpretation and Translation* 13:2. (2011a). 147-65.

Lee, C. “N-gram Analysis of Phrasal Differences between Translational and Original Korean News Texts.” *Interpreting and Translation Studies* 15. (2011b). 317-40.

Lee, C. “A Corpus-based Study of Simplification and Leveling-out in English-Korean Translation.” *Interpreting and Translation Studies* 17:3. (2013). 341-58.

Leuven-Zwart, K. M. “Translation and Original: Similarities and Dissimilarities, I.” *Target* 1:2. (1989). 151-81.

Leuven-Zwart, K. M. “Translation and Original: Similarities and Dissimilarities, II.” *Target* 2:1. (1990). 69-95.

Mahlberg, M. “Clusters, Key Clusters and Local Textual Functions in Dickens.” *Corpora* 2:1. (2007). 1-31.

Malmkjær, K. “Punctuation in Hans Christian Anderson's Stories and in Their Translation into English.” *Nonverbal Communication and Translation. New Perspectives and Challenges in Literature, Interpretation and the Media.* Ed. F. Payotas. Amsterdam: John Benjamins, 1997. 151-62.

May, R. “Sensible Elocution: How Translation Works In and Upon Punctuation.” *The Translator* 3:1. (1997). 1-20.

Nation, I. S. P. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press, 2001.

Nevalainen, S. “Köyhtyykö Kieli Käännettäessä? Mitätaajuuslistat Kertovat Suomennosten Sanastosta.” *Käännössuomeksi*. Eds. A. Mauranen & J. Jantunen. Tampere: Tampere University Press, 2005. 141-62.

Olohan, M. “Spelling Out the Optionals in Translation: A Corpus Study.” *UCREL Technical Papers* 13. (2001). 423-32.

Olohan, M. *Introducing Corpora in Translation Studies*. London: Routledge, 2004.

Olohan, M., & Baker, M. “Reporting *That* in Translated English: Evidence of or Subliminal Processes of Explicitation?” *Across Languages and Cultures* 1:2. (2000). 141-58.

Øverås, L. “In Search of the Third Code: An Investigation of Norms in Literary Translation.” *Meta: Translators' Journal* 43:4. (1998). 571-88.

Pastor, G. C., Mitkov, R., Afzal, N., & Pekar, V. “Translation Universals: Do They Exist? A Corpus-based NLP Study of Convergence and Simplification.” *Proceedings of the Eighth Conference of the Association for Machine Translation in the Americas (AMTA-08)* (2008). 75-81.

Puurtinen, T. “Explicitation of Clausal Relations: A Corpus-based Analysis of Clause Connectives in Translated and Non-translated Finnish Children’s Literature.” *Translation Universals: Do They Exist?* Ed. A. Mauranen & P. Kujamäki. Amsterdam: John Benjamins, 2004. 165-76.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. *A Comprehensive Grammar of the English Language.* London: Longman, 1985.

Scott, M. R. *Wordsmith: Software Tools for Windows.* Oxford: Oxford University Press, 1997.

Scott, M. N. *Normalisation and Readers’ Expectation: A Study of Literary Translation with Reference to Lispector’s A Hora da Estrela.* PhD Thesis, Liverpool: AELSU University of Liverpool, 1998.

Scott, M. *WordSmith Tools (Version 6)* [Computer Software]. Liverpool: Lexical Analysis Software, 2011.

Shlesinger, M. “Interpreter Latitude vs. Due Process. Simultaneous and Consecutive Interpretation in Multilingual Trials.” *Empirical Research in Translation and Intercultural Studies. Selected Papers of the TRANSIF Seminar, Savonlinna,*

1988. Ed. S. Tirkkonen-Condit. Tübingen: Gunter Narr, 1991. 147-55.

Stubbs, M. "Lexical Density: A Computational Technique and Some Findings." Ed. M. Coulthard. *Talking about Text. Studies Presented to David Brazil on His Retirement.* Birmingham: English Language Research, University of Birmingham, 1986. 27-42.

Stubbs, M. *Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture*. Oxford: Blackwell, 1996.

Toury, G. *Descriptive Translation Studies and Beyond.* Amsterdam & Philadelphia: John Benjamins, 1995.

Vanderauwera, R. *Dutch Novels Translated into English: The Transformation of a "Minority" Literature.* Amsterdam: Rodopi, 1985.

Xiao, R. "How Different is Translated Chinese from Native Chinese?: A Corpus-based Study of Translation Universals." *International Journal of Corpus Linguistics* 15:1. (2010). 5-35.

Xiao, R. "Word Clusters and Reformulation Markers in Chinese and English: Implications for Translation Universal Hypotheses." *Languages in Contrast* 11:2. (2011). 145-71.

Xiao, R., & Yue, M. "Using Corpora in Translation Studies: The State of the Art." *Contemporary Corpus Linguistics.* Ed. P. Baker. London: Continuum, 2009. 237-62.

Appendix 1

Make-up of the Comparable Corpus of American and Korean Novels (CCAKN)

A collection of 25 American novels (AN)

No	Title	Author	Publication date (year)	Release date on the website (year)	Size (tokens)
1	Adventures of Huckleberry Finn	Mark Twain	1884	2021	10,122
2	Babbitt	Sinclair Lewis	1922	2006	9,894
3	The Great Gatsby	F. Scott Fitzgerald	1925	2021	10,056
4	Life on the Mississippi	Mark Twain	1883	2006	9,961
5	Looking Backwards, 2000-1887	Edward Bellamy	1888	1996	10,024
6	Maggie: A Girl of the Streets	Stephen Crane	1893	1996	10,073
7	Main Street	Sinclair Lewis	1920	2006	9,984
8	McTeague	Frank Norris	1899	2006	9,975
9	My Antonia	Willa Cather	1918	2008	10,002
10	O Pioneers!	Willa Cather	1913	1992	9,996
11	Sister Carrie	Theodore Dreiser	1900	2011	10,036
12	Summer	Edith Wharton	1917	2018	10,155
13	The Age of Innocence	Edith Wharton	1920	1996	10,036
14	The Ambassadors	Henry James	1903	1996	9,988
15	The Awakening	Kate Chopin	1900	2006	10,031
16	The Call of the Wild	Jack London	1903	2008	9,973
17	The Education of Henry Adams	Henry Adams	1907	2011	10,020
18	The Jungle	Upton Sinclair	1906	2006	9,944
19	The Octopus	Frank Norris	1901	2008	10,024
20	The Portrait of a Lady	Henry James	1881	2001	10,003
21	The Rise of Silas Lapham	William Dean Howells	1885	2017	10,051
22	The Souls of Black Folk	W. E. B. Du Bois	1903	2021	10,048
23	The Turn of the Screw	Henry James	1898	2018	10,052
24	White Fang	Jack London	1906	2021	10,042
25	Winesburg, Ohio	Sherwood Anderson	1919	2012	10,003
	Total				250,493

A collection of 25 Korean novels translated into English (KN)

No	Title	Author/ Translator	Publication date (year)	Publication date of translation (year)	Size (tokens)
1	A Day in the Life of Kubo the Novelist	Tae-won Park/ Sun-young Park	1934	2010	10,033
2	A Distant and Beautiful Place	Kwi-ja Yang/ Chi-young Kim & Julie Pickering	1987	2002	10,000

3	Before and after Liberation	Tae-jun Yi/ Sun-young Park	1946	2010	10,008
4	Black and White Photographer	Yu-joo Han/ Janet Hong	2009	2009	9,129
5	Diary of a Vagabond	Yong Song/ Jason Park	1977	2008	10,103
6	Discovery of Solitude	Hee-kyung Eun/ Jae-won Edward Clung	2007	2008	10,012
7	Elephant	Jae-young Kim/ Mun-ok Lee, Nicholas Yohan Duvernay	2004	2008	9,871
8	Farmers	Mu-young Lee/ Young-nan Yu	1954	2002	10,209
9	From Wonso Pond	Kyung-ae Kang/ Samuel Perry	1934	2009	10,033
10	I Have the Right to Destroy Myself	Young-ha Kim/ Chi-young Kim	1996	2007	10,040
11	Land of Exile	Jung-rae Jo/ Marshall R. Pihl, Bruce & Ju-Chan Fulton	1981	2007	10,140
12	Lost Souls	Sun-won Hwang/ Bruce & Ju-Chan Fulton	1958	2010	10,039
13	Mujong	Kwang-su Yi/ Ann Sung-hi Lee	1918	2005	10,249
14	My Sister Bongsoon	Ji-young Gong/ Jung-eun Park	1998	2005	9,999
15	Raising Swallows	Dae-nyeong Yun/ Gabriel Sylvian	2007	2008	10,133
16	Sea and Butterfly	In-sook Kim/ Seok-ju Son, Catherine Torres	2003	2008	10,013
17	The Ancient Garden	Sok-yong Hwang/ Jay Oh	2000	2010	10,010
18	The Curse of Kim's Daughters	Kyong-ni Park/ Choon-won Kang, Myung-hee Lee, Kay-ho Lee & S.Keyron McDermot	1962	2004	10,067
19	The Dwarf	Se-hui Cho/ Bruce & Ju-chan Fulton	1978	2006	10,036
20	The Guest	Sok-yong Hwang/ Kyung-ja Chun & Maya West	2001	2005	10,116
21	The Korean Soldier	Sung-tae Jeon/ Jae-won Jung	2006	2008	10,171
22	The Man Who was Left as Nine Pairs of	Heung-gil Yun/ Marshall R. Pihl, Bruce	1977	2007	10,105

	Shoes	& Ju-Chan Fulton			
23	Three Generations	Sang-seop Yom/ Young-nan Yu	1932	2005	10,377
24	To Believe in Love	Yeo-sun Kwon/ Charles La Shure	2008	2009	9,365
25	Under the Fig Trees	Mi-kyung Jeong/ Ye-won Jeong	2004	2008	10,046
	Total				250,304

Appendix 2

List of the Top 50 Highly Frequent Lexical Bundles

	AN (Freq./ Range)			KN (Freq./ Range)		
1	one of the	108	24	in front of	117	25
2	there was a	94	22	out of the	108	24
3	it was a	93	22	one of the	62	19
4	out of the	78	22	in the middle	58	18
5	it was the	57	20	it was the	58	21
6	he was a	55	22	the first time	55	19
7	that he was	51	22	there was no	55	20
8	he did not	43	17	the middle of	52	20
9	he had been	43	20	a long time	51	18
10	that he had	43	21	it was a	50	20
11	there was no	42	15	there was a	45	19
12	part of the	37	19	the sound of	44	18
13	i want to	36	15	front of the	43	17
14	that it was	36	16	in the middle of	43	18
15	the end of	36	18	he had been	42	20
16	and in the	35	19	for the first	40	16
17	of the house	35	14	as soon as	39	17
18	that she was	33	12	that she had	39	11
19	to be a	33	18	the back of	38	17
20	would have been	31	14	the end of	38	18
21	he was not	30	15	as if he	37	14
22	here and there	30	16	be able to	37	18
23	to go to	30	14	in front of the	37	15
24	he would have	29	16	it would be	37	17
25	it was not	29	16	that i had	37	10
26	she had been	29	12	what are you	37	16
27	a sort of	28	14	for the first time	36	15
28	and there was	28	16	of the house	36	12
29	he could not	28	14	i don't know	35	18
30	some of the	28	14	seemed to be	35	17

31	the fact that	28	16	that it was	35	13
32	and it was	27	14	back to the	34	16
33	i was a	27	15	do you think	34	18
34	it would be	27	14	was about to	34	15
35	one of them	27	15	but it was	33	17
36	when he was	27	14	for a while	33	14
37	but it was	26	16	middle of the	33	17
38	end of the	26	13	that he was	33	17
39	for a moment	26	12	the middle of the	33	17
40	in front of	26	14	i want to	32	13
41	the edge of	26	12	that he had	32	15
42	to do with	26	14	a few days	31	16
43	in the world	25	15	for a long	31	17
44	side of the	25	14	the rest of	31	18
45	up and down	25	12	next to the	30	13
46	and he was	24	13	on the floor	30	14
47	on the other	24	16	to be a	30	17
48	was going to	24	10	for a long time	29	16
49	which he had	24	13	if he were	29	14
50	back to the	23	13	down on the	28	17

[Abstract]

A Corpus-based Study of Translation Universals in Korean-English Literary Translation

Woonhyung Chung
(Yonsei University)

Translation universals hypothesizes that translational languages have conventional yet unique features of translations, distinct from those of non-translations, and these features are universally present in all translations regardless of language pairs. However, the verification of translation universals has been restricted mainly to genetically close language pairs, such as English and closely related European languages. To confirm its *universality*, this article investigated Korean-into-English translations in the literary genre, which have remained mostly unexplored in translation universals studies. By constructing an approximately 500,000-word comparable corpus—the Comparable Corpus of American and Korean Novels—this study tested four major sub-hypotheses: simplification, explicitation, normalization, and leveling-out. Setting aside the lexical density of simplification indices, the results from all the other indices—mean sentence length, proportions of high- and low-frequency words, normalized frequencies of connectives and lexical bundles, and standard deviations of lexical density and mean sentence length—supported the translation universals hypothesis, indicating its high generalizability.

Key words: translation universals, Korean-to-English translation, literary translation, corpus-based translation studies, comparable corpora

Woonhyung Chung
Yonsei University, Doctoral Student.
Department of English Linguistics and Language, Yonsei University,
50, Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea
02-2123-2300
E-mail: woonka82@hanmail.net

접 수 일: 2021년 12월 31일
심사기간: 2022년 1월 1일~2월 18일
게재결정: 2022년 4월 5일