

The hallmarks of L2 writing viewed through the prism of translation universals*

Younghee Cheri Lee
(Yonsei University)

Lee, Younghee Cheri. 2018. The hallmarks of L2 writing viewed through the prism of translation universals. *Linguistic Research* 35(Special Edition), 171-205. Rooted in a perception that second language (L2) writing bears neither resemblance with nontranslated counterparts nor relation to translation, this article explores the untested terrain of revealing lexical and textual attributes unique to L2 writers' texts, thus identifying their linguistic qualities from the angle of translation universals. Setting plausible parameters to discern translational instances related to lexical and syntactical choices, this article argues that idiosyncratic properties shared by translated English may typify the hallmarks of L2 writing produced by non-anglophone scholars in English disciplines. By compiling the comparable corpora of English journal abstracts consisting of 638,764 tokens, it is shown how salient translational features arise in expert L2 writers' texts in compliance with corpus linguistics. Kruskal-Wallis tests are applied to evaluate linguistic indices that make Korean scholars' L2 writing distinct from native scholars' original writing. On a substantial level, a general presumption on the interrelatedness between expert L2 writers' English and translational English has turned out to be warranted, meaning that Korean scholars' L2 writing can be marked by universals of simplification, normalization, explicitation, and convergence in their broad outlines. It can be deducible from the findings that regardless of L2 proficiency levels, second language writers may be destined to go through a 'mental translation' as an inescapable cognitive mechanism during the L2 writing process, which in turn renders translational manifestations pervasive in the 'product' of L2 writing. The terminal pedagogical aim building metacognitive awareness to be mindful of second language processing, thus, this article concludes that expert L2 writers need to equip themselves with metacognitive strategies, thereby being consciously and explicitly aware of what to avoid and what to accept during the process of L2 writing entailing mental translation. (Yonsei University)

Keywords second language (L2) writing, L2 writers' texts, translation universals, mental translation, comparable corpora, corpus linguistics, journal abstract

* This article is adapted from the author's unpublished doctoral dissertation (Lee 2017). The author gratefully acknowledges the editor and the two anonymous reviewers of *Linguistic Research* for their thoughtful comments and suggestions for improvement. All errors remain the author's own.

1. Introduction

Since the computer frontier became one of the primary aspects of human life, the computer as a transactive memory partner has influenced the way people carry out their life agendas in many different aspects. Concurrently, since the 1980s, when cutting-edge computer technologies entailing natural language processing algorithms significantly accelerated the digitization of texts, the advent of corpus linguistics has facilitated researchers' inquiries in scholarly disciplines, thus enabling them to pursue their goals from different perspectives. In effect, where any second language-related disciplines are concerned, it is indisputable that researchers often face the problem of not having linguistic intuitions in a second language. In virtue of corpus linguistics, many non-native linguists and non-anglophone scholars, even without intuition or introspection, have been motivated to devote themselves to the study of a second or foreign language, and this field eventually attracted significant scholarly attention.

Although research on L2 studies has advanced recently, there remain unmet research needs for further exploration, one of which would be the lack of studies on second language writing with an approach to corpus linguistics. In the world of 21st-century skills that are of paramount importance, 'writing' as a cognitive communication skill grabs full attention as one of the core competencies, and corpus-based L2 writing studies have substantial empirical research prospects. Nonetheless, only limited attempts have been made to uncover the linguistics nature and textual properties of L2 writing (Cumming 2012; Kaplan 2000). The case seems no different in the domestic research context for two reasons. First, L2 writing research to date has centered primarily on delineating the features of unskilled writers, rather than those of skilled writers (Chae 2012; Kim 2010). While a growing number of studies have raised awareness of unskilled Korean L2 writers, previous studies are not necessarily extendable outside of the research context in that they still lack a coherent understanding of the linguistic properties that typify Korean L2 writing in a broader sense. Moreover, L2 writers' texts have not been the focus of the research to discuss textual attributes from the angle of translation universals. The ultimate aim is thus to compare them with nontranslated original L1 writing, thereby perceiving linguistic properties through the prism of translation-related features if any.

As a first step to fill in the two missing holes in the previous research, it seems noteworthy to be aware of the relationship between the two languages. As Cumming (1990) claimed, L2 writing is a 'bilingual event' associated with multiple cognitive strategies. L2 writing is distinct from L1 writing in that L2 writers incorporate the use of L1 in the process of L2 writing (Leki et al. 2010; Silva and Brice 2004; Woodall 2002). As regards the interrelatedness between the two languages, it was originally claimed by Cook (1992: 558) that "the L1 and L2 languages share the same mental lexicon so that L2 processing cannot be cut off from L1." On the premise that second language writers manipulate both languages interchangeably as a natural consequence of this inescapable cognitive process of L2 writing, the properties of L2 written production would be neither those of authentic English writing nor those of entirely translated English writing. Given that L2 writers produce texts that differ predictably from baseline nontranslated texts, then, what are the distinctive linguistic qualities that these two different texts may hold? If the linguistic properties drawn from L2 writers' texts are qualitatively similar to those of translated texts, how are the shared features related to each other? With these research queries in mind, this article works towards uncovering the distinctive linguistic qualities of expert L2 writers' texts. No matter how intertwined L2 writers' texts and translated texts may be, these empirical research efforts would render our knowledge of Korean L2 writers' texts more precise, hopefully enabling us to eventually determine what predictors and indicators might be more valid and accessible to detect translational instances in L2 writers' texts. The application of corpus-based descriptive translation studies, along with computational tools and robust algorithmic treatments, would facilitate this mindful endeavor to identify deeper-level linguistic attributes inherent to expert L2 writers' texts.

Taking the research aims into account, the discussion above leads to the following research questions:

- What features viewed through the perspective of translational universals make Korean L2 writing differ from native speakers' original writing?
- Concerning the hypotheses of translational universals, what are the linguistic indices that typify Korean L2 writers' texts?

2. Literature review

2.1 The cognitive process of L2 writing

The interrelatedness between L1 and L2 writing has long been the locus of L2 writing research. The early L2 writing model was associated with the cognitive process of L1 writing (Flower and Hayes 1981; Grabe and Kaplan 1996; Kellogg 1987, 1999; Schoonen et al. 2003; Wang and Wen 2002). The first writing model was put forward by Flower and Hayes (1981), by entailing the acceptance that L1 and L2 writing might be closely intertwined with each other. These researchers challenged a shift from the conventional models of the writing process, thus proposing a basic cognitive writing model involving cognitive roles during the recursive cycle of writing. Included in their first model were three writing phases: planning, translating and reviewing. Despite the cognitive process of L1 writing, however, this model only accounted for the course of writing in one's mother tongue. To further augment the operation of the revision phase of the original model, they later on adjusted their initial version by incorporating the two phases of 'writer processes' and 'writer knowledge,' aiming to depict more specific cognitive paths entailed during the phases of evaluation and revision. While Flower and Hayes did not pay much attention to translation as the primary event of the writing process, Kellogg's model gained attention as it incorporated translation as the primary phase of writing event preceded by planning during the formulation phase (Grabe 2001).

The L2 writing model produced by Wang and Wen (2002) is the only available model that has been meticulously designed to account for the cognitive process of L2 writing more thoroughly than traditional models. Even though Flower and Hayes' model has induced the factor of cognition into writing, its treatment seemed to remain relatively minimal since their model only pertains to L1 writers, not extending the reach of the L2 writers. To better represent the association of the two languages manipulated in the process of writing, they emphasized depicting the interrelated and discrete functions of the two languages. Overall, it appears that mental translation has been a primary event of L2 composition that L2 users are destined to go through. Nevertheless, there remain little attempts to identify whether or not a mode of mental translation affects L2 written products in the aspect of linguistic quality.

2.2 Translation as an L2 writing strategy

L2 writers at various proficiency levels have applied different types of writing strategies in the L2 writing process. In particular, translation has been applicable as a viable strategy with different purposes across all ranges of L2 writing proficiency. Uzawa (1996) acknowledged that using translation as a writing strategy benefits writers at lower levels of writing proficiency, thus enabling these writers to think, transform, organize and present their ideas in L2 by translating L1 texts. Kim (2009) claimed translation could be considered a product generated by the use of the syntactic, rhetorical and lexical features of L1 during the L2 writing process. Gosden (1996) asserted that less-skilled writers at the tertiary level favored writing in L1 first and then move on to employ a phrase-by-phrase translation strategy, rather than to generate their writing in L2 directly. Regarding writing fluency along with a translation strategy and pausing behaviors throughout the L2 composition process, Sasaki (2000) conducted a large body of research comparing experts with novices and comparing skilled writers, less-skilled writers, and novice writers, respectively. Evidence indicated that pausing behaviors occurred in both competent and incompetent writers but for different purposes. Incompetent L2 writers applied a pausing strategy to translate their thoughts into L2, whereas more competent L2 writers paused to refine L2 expressions in their written products (Bagheri and Fazel 2011). Wang and Wen (2002) proposed that skilled L2 writers might be less inclined to employ the strategy of translation but to write directly in the target language, whereas less-skilled L2 writers were apt to count on translation as an L2 writing strategy. Overall, it is worthwhile to note that these researchers assumed that translation involves not only transforming L1 scripts into the target language but also searching resources of the target language using their L1.

2.3 L2 writers' texts

The past research endeavors regarding L2 writers' texts were confined to dealing with a small scale of data; thus, few studies have incorporated corpus-based analysis. In recent years, the research conditions have been so

much improved compared to the past decade, but there still need to forge new pathways, by adding more varieties to corpora and developing more outperforming corpus-analyzing tools. L2 writers' texts are typically studied from the two-tier perspectives: one is concerned with linguistic features of L2 texts while the other pertains to rhetorical differences between L1 and L2 texts. Some researchers (Connor 1990; Swales 1990) compared L2 writing produced by writers at different levels with a different L1 background in order to discover how L1 affected L2 writing concerning the better behaviors of lexical and rhetorical choices and found out a positive interplay between the use of L1 and L2. Regarding the use of cohesive devices, Qi (1986) observed that L2 writers were firmly attached to the frequent use of cohesive devices such as conjunctions. Hinkel (2002) provided the most precise information about linguistic properties prevalent in L2 texts using a corpus of 434,768 tokens by comparing native speakers' original English texts to L2 writers' texts from different non-Anglophone countries such as Japan, Korea, China, Vietnam, Indonesia, and Arabia. Hinkel asserted that L2 writers' texts hold the features of syntactic and lexical simplification. Studies on L2 scholarly journal abstracts produced by Korean scholars are only limited to discuss the 'move' structures with an approach to rhetorical analysis. Overall, there has been a consistency in the findings of such research endeavors, most of which concerns the overuse of additive conjunctions and linguistic sophistication.

2.4 Translation universals in L2 writers' texts

The origin of the textual study goes back to the late 50s when it was even before the advent of corpora. In the early 1990s, research techniques with electronic corpora in Translation Studies were primarily initiated by Baker (1993), which started to gain attention as substantial empirical evidence from a range of corpus-based research into translation universals (Baker 1996; Laviosa 1998a).¹

1 Translation universals (Baker 1993) have been raised by many scholars with different notions: the third code (Frawley 1984), hypotheses (Blum-Kulka 1986), translationese (Gellerstam 1986, 1996), laws (Toury 1995; Chesterman 2004a), core patterns (Laviosa 1998b), features of translation (Olohan 2004), universals (Baker 1993; Chesterman 2004b), regularities of translations (Zanettin 2012), universals in language typology (Mauranen 2007), skepticism on translation universals (Tymoczko 1998; Malmkjær 2007; House 2008, 2012), translational behavior (Toury 1995; Malmkjær

Such studies highlight that translation is neither target language nor source language, in the sense that translational language bears its peculiar linguistic properties produced by both languages. Baker (1993) defines that translation universals typically occur in translated texts rather than original utterances by being the inevitable by-products of the manipulating process rather than the effects of L1 interference. She also claims that specific patterns that are apparent in all sets of translated versus nontranslated texts would suggest a hypothesis for universal features of translation (Baker 1995). Along the same line, Laviosa (2002) supports that these features are recurrent linguistic properties shared in all translated texts, which differ not only from their source texts but also from comparable texts in the target language. The recent approach to descriptive translation studies has defined translation as a product, by comparing translated corpora to the corpora of nontranslated native texts. Reflecting that translation universals are cognitive phenomena, Baker (1993) argues that all translations are apt to hold specific linguistic characteristics only by being translations, which are caused in and by the process of translation. Research into the characteristics of the translational language is considered beneficial as it promotes linguists and translation scholars to a higher level of understanding on the nature of translation. Concurrently, it brings writers' language awareness of the conscious or unconscious cognitive processing involved in the direct or indirect translational activity (Chesterman 2004a, 2006).²

Corpus-based descriptive translation studies have primarily concerned a variety of the translation universals hypotheses, the best known of which comprises the four potential universals proposed by Baker (1996): the simplification, normalization, explicitation, and convergence hypotheses.³ Simplification has been the locus of the research into translation universals involving several predictors to gauge the instances of translational simplification,

2012), unavoidable effect of translationese (Hartmann 1985; Baker 1993; Teubert 1996; Laviosa 1997; McEnery and Wilson 2001; McEnery and Xiao 2002, 2007), unrepresentative special variant of the target language (McEnery and Xiao 2008).

2 Chesterman's proposals for translation universals include: (1) the relation between translations and their source texts, and (2) the association between translations and comparable non-translations in the target language (2004b).

3 Empirical research findings regarding the sub-hypotheses of translation universals appear in more detail in the author's doctoral dissertation (Lee 2017).

which refers to the tendency to simplify translational language lexically, syntactically, and stylistically. Such technique to simplify the content and the language in translated texts *per se* attributes to the translators' purposeful intention to adjust the level of readability. Translated texts are expected to show lower levels of lexical variety, sophistication, richness, and density than those of nontranslated texts (Baker 1996). Some valid parameters to discern lexical simplification include Standardized Type/Token Ratio (STTR), function words versus lexical words, top-to-bottom frequency words (Laviosa 1998b, 2002).

Normalization centers on the notion that the atypical or unattested language of translation is more apparent in target texts than source texts, thus causing somewhat awkwardness in translations. Baker (1996: 183) defines normalization as the "tendency to conform to patterns and practices that are typical of the target language, even to the point of exaggeration." The amount of lexical bundles in translated texts is one of the most commonly used predictors to test normalization. Translated texts are thought to contain a higher portion of lexical bundles. Typical manifestations of normalization also include overuse of clichés or common grammatical structures of the target language and overuse of typical features of the genres involved.

Blum-Kulka (1986) observed the evidence of explicitation through a few individual sample texts by which she could uncover translators' behavior of making translated texts more explicit and concrete. Translators exploited some strategies similar to adding extra cohesive devices such as connectives, regardless of prevailing redundancy. These linguistic counterparts were not explicitly present but implicitly connoted in the source text, though (Øverås 1998). Such translation technique of explicitation is, as part of syntactical explicitation, aimed at increasing the clarity of information delivered from the source to the target text. Baker (1996: 180) also argues on explicitation in a broader sense that its occurrence is closely linked to the techniques or strategies to "spell things out rather than leave them implicit." For instance, explicitation occurs when conjunctions are used more frequently in translated texts than in nontranslated texts. Explicitation has been the most investigated and the least controversial TU hypothesis to date (Xiao and Dai 2014).

The universals of leveling out (also called ‘convergence’) relates to the “tendency of translated text to gravitate towards the center of a continuum” (Baker 1996: 184), thus translators as L2 writers typically steering a middle course between extremes. As a consequence, this affects translated texts to be more similar to one another than their nontranslated counterparts. Laviosa (2002) confirms that translated texts show a relatively higher level of homogeneity. Along the same vein, Olohan (2004) also regards the universals of convergence as a tendency of less variance of translated texts than their counterparts. Some examples of convergence predictors include readability indices, standards deviations of mean sentence length and the STTR values (Goh and Lee 2016).

3. Research method

3.1 Corpus design

In an attempt to attain the proposed research goals, compiling the purpose-end Comparable Corpora of English Research Abstracts of Scholarly Journal Articles (henceforth, CCERA) was of vital necessity for two motivations. The first necessity was driven by the unavailability of non-literary corpora of translated English.⁴ Seeing that the TEC may entail little external validity outside research contexts, especially involving language pairs with genetic distances, the other necessity has come from the inaccessibility of comparable corpora produced by expert writers of English, whose L1 and L2 are genetically unrelated.

The CCERA as monolingual corpora includes 2,243 English research abstracts in linguistics and English literature to compare and contrast nontranslated native English with two types of Korean scholars’ English abstracts whose author/s are not affiliated with Anglophone institutions, thus the fundamental aim conducting a three-way comparison in each domain respectively.

4 Although the Translational English Corpus (TEC; Baker 2004) is the only publicly available contemporary corpus of translated English produced by native speakers, due to its scale and genre restrictions, most prominent research concerning translational English has been mostly constrained to uncovering the textual properties and registers of literary texts, especially fiction.

In each discipline, the subcorpus of English abstracts based on Korean articles was named *translated* to represent translated L2 writing in the supposition that Korean scholars' English abstracts may have used their journal articles written in Korean as a source text (henceforth, the KE subcorpus). Another subcorpus of English abstracts based on English articles was labeled as *quasi-translated* to represent quasi-translated L2 writing in the speculation that their English abstracts are neither based on the L1 source language nor formal translation (henceforth, the EE subcorpus). The third subcorpus of American scholars' English abstracts was titled *nontranslated* to represent L1 writing (henceforth, the AE subcorpus). By having comparable sets of data (nontranslated versus quasi-translated, nontranslated versus translated, and quasi-translated versus translated English abstracts) across the disciplines, in turn, we will further discuss whether the linguistic tendencies and instances traced are domain-dependent or universal.

The six subcorpora of the CCERA were compiled to be as balanced as possible for a more equitable comparison, especially concerning size, a time span of production, genre representation, and search terms regarding specific sub-topic areas and registers. The time span taken into consideration covers from 2000 to 2016. The texts for the CCERA corpora were selected from numerous databases with high accessibility to a vast number of prestigious international scientific journals as well as acclaimed domestic journals.⁵

The abstracts for the CCERA were carefully selected to represent each category's identity: *nontranslated* L1 writing versus *quasi-translated* L2 writing versus *translated* L2 writing. Based on the method of simple random sampling, individual abstracts were chosen at random and not more than once to prevent any bias that may negatively affect the validity of research findings.

As for the compiling process of the CCERA, the electronic versions of the journals were accessed, and the files were converted and saved in text format with their title placed in angle brackets on top of each abstract compiled. The scale of the CCERA is mapped out in Table 1.

5 The databases accessed on necessity include RISS-International DB, Proquest Literature Online (LION), Linguistics and Language Behavior Abstracts (LLBA), Directory of Open Access Journals (DOJA), Academic Search Complete (EBSCOhost), eARTICLE, and DBPIA.

Table 1. The scale of the CCERA corpora⁶

Domain	Corpus	Text Feature	Abstract	Token	Type	Average TextLength
LINGUISTICS	AE_LING	nontranslated (NT)	600	105,535	7,594	176
	EE_LING	quasi-translated (QT)	605	106,195	6,139	176
	KE_LING	translated (TT)	603	106,545	5,898	177
	Sub Total		1,808	318,275	11,473	
ENGLISH LITERATURE	AE_LIT	nontranslated (NT)	530	106,851	9,743	202
	EE_LIT	quasi-translated (QT)	440	107,869	8,538	245
	KE_LIT	translated (TT)	435	105,769	9,086	435
	Sub Total		1,405	320,489	16,450	
Total			2,243	638,764		

3.2 Data analyses

Delineating the instances of translational universals across the CCERA, this study carried out two quantitative methods of descriptive and inferential statistics. For the baseline analysis, numerical values on each hypothesis were computed automatically using two corpus analyzing tools: WordSmith Tools 7.0 (Scott 2013) and AntConc 3.4.4w (Lawrence 2013). Further analysis for inferential statistics was proceeded to provide the central tendency and dispersion observed in each subcorpus of the two disciplines. From the baseline analysis of normality tests, it was observable that the data did not follow a normal/standard distribution. The Kruskal-Wallis test (the non-parametric version of one-way ANOVAs) was applied throughout the analyses accordingly.⁷ The Pairwise Wilcoxon test was adopted as a post hoc test to evaluate the statistical significance of mean differences among paired corpora.

⁶ Specific sources and descriptions of the data compiled for the CCERA can be found in the author's dissertation (Lee 2017).

⁷ Following the procedures suggested by Gravetter and Wallnau (2013), the data collected for this study were normalized using a 'z-score' transformation. As the z-scores based on the STTR values (frequency) were different from those grounded on connectives (percentage), the z-scores for each linguistic factor were then converted into normalized values. Note that the transformed value 22.74 is more intuitive than its z-score -0.529 for STTR.

Finally, to further augment critical findings and to detect additional instances of translational manifestations, a qualitative analysis was additionally performed by investigating into regular and irregular linguistic patterns and the tendencies of individual items within the extended context so that it can compensate for insufficient explanations. Items put into the qualitative analysis were mostly those that ranked high in the frequency analysis or those that showed meaningful or exceptional patterns of TU instances so that the different natures of those items could be in comparison in quality.

3.3 Measurement variables

Concerning the hypotheses of translation universals postulated by Baker (1996) and Laviosa (1998a), this study applied six primary linguistic indices to trace the four types of translational instances, which include translation universals of simplification, normalization, explicitation, and convergence hypotheses.

First, lexical simplification was detectable by three measurement variables. The STTR index was used to trace lexical variety, bottom-frequency words for lexical sophistication, and top-frequency words for lexical richness. The first one gauges the lexical variety of the translational English of Korean scholars employing the STTR scores.⁸ If Korean writers' L2 texts show evidence of lower lexical diversity, they may have a narrow range of vocabulary, coinciding with the premise that L2 writers might have recycled the same vocabulary repeatedly across the texts.

Translated texts are expected to show a more significant preference for high-frequency words and less favoritism towards low-frequency words than those in nontranslated texts. Regarding lexical richness as a second indicator, the top-20 recurrent words were compared.

8 The STTR is often considered to complement the type/token ratio (TTR) in that the TTR could be a less plausible index, especially when the gap of the length of the two texts in comparison is too high (Scott 2004). The STTR represents the average of every 1,000 words' TTR of the texts compared, thus enabling valid comparison among multiple texts with different lengths. As the length of most research abstracts does not normally exceed 500 words, however, this study computed the STTR values based on every ten words' type/token ratio of the six subcorpora in the CCERA, thereby ensuring valid comparison.

As a third indicator for lexical simplification, the degree of lexical sophistication was measured by applying the low-frequency profile across the CCERA. The low-frequency profile is estimated by the portion of bottom-frequency words which is based on the words with only one-time occurrence among the sub-corpora in both academic fields.

Also, to discern the pattern of translation universals of lexical normalization, the numerical values of lexical bundles were measured by using the N-gram in the AntConc software program. The N-gram size was set to three, while the minimum frequency was set to two due to the constraints of text length. The proposed explicitation hypothesis premises that there would be more evidence of cohesive linguistic devices in translated texts than their counterparts.

Then, to investigate the universals of syntactic explicitation, the frequency of connectives was counted across the CCERA to figure out how Korean scholars' texts increase cohesive relationship. As the sub-hypothesis of translational explicitation assumes, testing the number of connectives was the focus so that the values of each corpus were measured accordingly using the reproduced list of 114 connectives based on the original list of 105 connectives provided by Quirk et al. (1985) and Biber et al. (1999).

Finally, to discover the instances of translational convergence, the level of homogeneity was measured by observing how Korean scholars' texts are analogous to each other. As the convergence hypothesis posits, translated texts are inclined to converge at the center along the continuum, which renders their high homogeneity having less variety in between the texts. The standard deviations of mean sentence length were computed using the WordSmith Tools 7.0. In a general view of the hypothesis of convergence, translated texts are expected to show a lower standard deviation in mean sentence length, thus indicating that translated texts are more homogeneous in a syntactic aspect than nontranslated texts.

4. Results⁹

4.1 Universals of lexical simplification

4.1.1 Lexical variety: STTR

As a measurement variable of lexical simplification, lexical variety employed the STTR values to judge how many different types of words occur in a given text. Like said, the extent to which corpus group holds a higher diversity of vocabulary was evaluated to determine whether L2 writings are lexically more simplified than their counterparts.

During the baseline analysis by descriptive statistics, the STTR figures in each subcorpus were lined up in descending order, and both domains showed quite a similar pattern (see Table 4). The mean ranks of each subcorpus in the domain of linguistics were listed in descending order from the nontranslated to the quasi-translated, then to the translated corpora, resulting in statistical significance ($\chi^2(2)=96.008$, $p<.001$). Further observation by post-hoc tests demonstrated that mean differences were statistically significant across the linguistics (AE-EE: $p<.001$, AE-KE: $p<.001$, EE-KE: $p<.001$).

In English literature, the mean values of each subcorpus came out in descending order. The mean values were found to be statistically significant ($\chi^2(2)=36.999$, $p<.001$). The mean differences also reported statistical significance, except for the pair of the quasi-translated versus translated corpora (AE-EE: $p<.001$, AE-KE: $p<.001$, EE-KE: $p=.150$). Detected by the variable of STTR, results reported that expert L2 writers' translated and quasi-translated English texts tend to be lexically simplified. The STTR values were evident to a lesser degree in Korean scholars' L2 abstracts, which are, as expected, restricted in their use of varied vocabulary, thus showing lower lexical variation than nontranslated original texts.

9 Note that the results of statistical analyses for each variable are comprehensively summarized in Table 4 at the end of the Results section.

4.1.2 Lexical sophistication: bottom-frequency words

Lexical sophistication employed bottom-frequency words as its parameter. Similar to lexical variety, lexical sophistication is indicative of L2 writing proficiency. The only difference would be that lexical sophistication is based on the proportion of unusual and comparatively advanced words. Based on the descriptive statistics, the linguistics corpora appeared to have comparatively smaller portions of low-frequency words than the other discipline. As expected, the mean ranks of each subcorpus in linguistics lined up in descending order from nontranslated to quasi-translated, then to translated corpora with statistical significance ($\chi^2(2)=146.970$, $p<.001$). The mean differences of all the paired corpora in the linguistics domain were found to be statistically significant (AE-EE: $p<.001$, AE-KE: $p<.001$, EE-KE: $p=.032$).

In English literature, the mean values of each subcorpus showed the similar pattern as the case in linguistics, listed up high to low from nontranslated to translated with statistical significance ($\chi^2(2)=112.930$, $p<.001$). The post-hoc tests reported that the nontranslated corpus had a more considerable portion of bottom-frequency words than its counterparts, and mean differences were statistically significant (AE-EE: $p<.001$, AE-KE: $p<.001$). The quasi-translated corpus had a lower portion of bottom-frequency words than the translated corpus. Their mean difference was statistically insignificant, suggesting a possibility to move towards the shift of lexical simplification (EE-KE: $p=.460$).

4.1.3 Lexical richness: top-frequency words

The frequency level of recurrent words in translated L2 texts demonstrates whether the texts are lexically simplified. The larger the type of top-frequency words but the lower the values of them, the more vibrant the words a text represents, and vice versa. On the same continuum as lexical variety, lexical richness was also measurable based on the number of types running in a corpus, but the only difference was that lexical richness stands by frequency - that is, the rate at which a type is recurrent in a given corpus.

Drawn by descriptive statistics, the proportion of top 20-frequency words

showed sequential order. The KE_LING corpus had the highest portion of them while the AE_LIT corpus has the least. Also, it was further observable that the mean ranks of each subcorpus in the linguistics corpora showed ascending order from the nontranslated corpus to the quasi-translated corpus, then to the translated corpus, resulting in statistical significance ($\chi^2(2)=142.730$, $p<.001$). Followed by the post-hoc comparison test, the mean differences of top-frequency words were turned out to be statistically significant as well (AE-KE: $p<.001$, AE-KE: $p<.001$, EE-KE: $p<.001$). In the domain of English literature, the mean ranks of each subcorpus showed ascending order from nontranslated to quasi-translated, then to translated KE_LIT corpora, showing statistical significance ($\chi^2(2)=19.194$, $p<.001$). Further, post-hoc tests reported that mean differences involving the nontranslated pairs were statistically significant (AE-EE: $p<.001$, AE-KE: $p<.001$). However, the same result was not visible between quasi-translated and translated corpora (EE-KE: $p=.750$).

Overall, both the linguistics and English literature corpora were found to hold a higher portion of top-frequency words corresponding to the universals of lexical simplification. Concerning the linguistic features of the quasi-translated texts, the results of both domains were not parallel to each other as the gap between the EE_LING and KE_LING pair was valid. Likewise, the same pattern was not apparent in the other discipline. The EE_LIT and KE_LIT pair showed the reversed order of numerical values and showed statistical insignificance in group differences, thus being consistent with lexical simplification.

As part of the qualitative analysis, top-frequency words were looked through manually to observe a significant linguistic tendency pertinent to the instances of lexical richness, and such a case was detected. To be specific, based on the 100 most frequently recurring vocabulary words, only the verbs and nouns, which directly relate to abstract writing as a genre, have been given particular attention, and noteworthy results were detectable. As for the composition of verbs and nouns for journal abstracts, the nontranslated AE_LIT and AE_LING corpora demonstrated the highest level of lexical richness, with the most varied types of nouns and verbs pervading across the corpora with almost equal frequencies. Specifically, when stating the purpose of research and developing arguments, as well as manipulating reporting verbs, native writers manipulated more varied types than their counterparts in both disciplines. Besides, the

evidence that native writers make the same lexical choices as Korean scholars was not detectable. For example, only the nontranslated corpora showed a high preference for the noun *article*, which took up about 0.05% in Korean scholars' corpora (see Figure 1). Conversely, there were two nouns that Korean writers firmly attached to them while their counterparts did not. The nouns *purpose* and *aspect* demonstrated such a case, especially in the KE linguistics corpus. Moreover, the noun *study* was always preferred to *paper* across the CCERA, except for the quasi-translated and translated corpora in English literature.

The case with verbs was not much different from that with nouns. All the nontranslated corpora contained the most abundant types of verbs, recycling in a balanced way, but not showing favoritism towards particular verbs. To be specific, the verb *argue* was noticeable as it seemed to be the first option for native writers but not for Korean writers. Instead, Korean scholars in linguistics showed a particular preference for *investigate* and *examine*. The same lexical composition was not observable in the other domain, in any case. None of the writers in English literature were inclined to *investigate*, ranked below the 500th in English literature.

As a domain-dependent observation, some meaningful instances were discernable. The linguistics corpora always formed 'overlapping' circles with one or two 'intersections,' thereby the nontranslated AE corpus never becoming the 'union' of Korean L2 writers' corpora. Based on the observation, it seems plausible to predict that the intersections composed of translated and quasi-translated writing may be the evidence of translational manifestations inherent to Korean L2 writers' texts. In contrast, the English literature domain constructed three 'concentric' circles by having the nontranslated corpus in the fringe, the quasi-translated corpus in the frontier, and the translated corpus in the core whether in case of nouns or verbs. Figure 1 illustrates the detailed output.

When taking the writer's stance or identity into account, lexical richness seems to be of the utmost significance, and more importantly, the language use behavior as such can help writers eliminate 'translationese' from their writing. In a broader context, Hyland (2002c) also asserted the use of different reporting verbs by varying the strength of writers' voices depending on discipline shows the importance of their voice as an author writing in the target language.

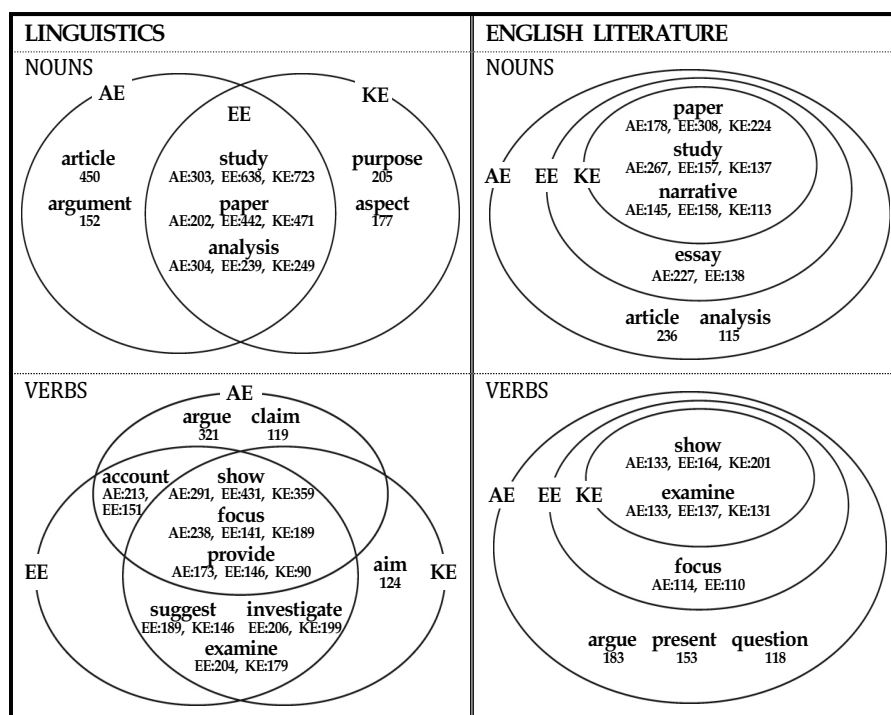


Figure 1. The distinctive compositions of overlapping nouns and verbs

4.2 Universals of lexical normalization: lexical bundles

The statistical averages of highly-recurring lexical bundles across the linguistics corpora were lined up in ascending order. In linguistics, the translated corpus contained a more substantial number of lexical bundles than nontranslated versions. Among the six subcorpora, the KE_LING corpus turned out to be the highest lexical bundle holder, containing even more lexical bundles than the KE_LIT group. The same trends were not detectable across the other discipline, in any case. The stated consistency result did not hold, as the statistical values of lexical bundles in English Literature were starkly opposed to those of linguistics. The nontranslated corpus in linguistics recorded lower values than their counterparts, whereas the AE_LIT corpus showed higher values than the other two in English literature. The KE_LIT corpus had the highest portion

of lexical bundles whereas the EE-LIT corpus had the lowest. Further analyzed in linguistics, the mean ranks of each subcorpus demonstrated statistical significance ($\chi^2(2)=99.797$, $p<.001$). The lexical bundles of nontranslated English texts were significantly higher than those of quasi-translated and translated corpora (AE-EE: $p<.001$, AE-KE: $p<.001$, EE-KE: $p<.001$). In English literature, the mean values of each subcorpus reported marginally significant results ($\chi^2(2)=5.149$, $p=.076$). The post-hoc tests showed that the mean differences of lexical bundles were all insignificant across the subcorpora pairs (AE-EE: $p=.140$, AE-KE: $p=.140$, EE-KE: $p=.780$). Although the mean ranks of every subcorpus in English literature resulted in the reversed order of the hypothesis, the post-hoc proved them to be statistically insignificant, which thus supports the hypothesis. Overall, both disciplines were found to have the values of lexical bundles corresponding to the hypothesis of normalization, being compatible with translation universals. Concurrently, the findings indicate that lexical bundles as a translational parameter may well account for the tendency of lexical normalization.

An additional observation was made to evaluate writers' linguistic behaviors qualitatively. The composition of high-frequency lexical bundles in the linguistics corpora was examined using the 3-gram lexical bundles ranked up to top 50. As demonstrated in Table 2, it was recognizable that each subcorpus formed its distinctive combination of lexical bundles so that recurring lexical bundles showed a significant discrepancy especially in the domain of linguistics. For example, the 3-gram lexical bundle *it is argued* was the most highly recycled phrase in the AE_LING corpus, *the effect of* and *the purpose of* in the EE_LING corpus, and *the purpose of* in the KE_LING corpus. As expected, the most varied types of verb phrases were considerably marked in the nontranslated corpus and the least different kinds in the translated corpus, whereas the least distinctive *noun-of* bundles were notable in the nontranslated corpus and the most varied ones appeared in the translated corpus. As the composition of lexical verb phrases contained in the translated corpus was distinct from that of their counterparts, thus, it seems reasonable to conclude that the extensive use of *noun-of* bundles would be one of the most robust translational features unique to Korean scholars' texts.

Concerning the use of lexical bundles, there are specific types of set phrases preferred in journal abstract writing. According to the Publication Manual of the

American Psychological Association (2010), writers are encouraged to use ‘verbs’ rather than their ‘noun’ equivalents. It is accepted as the standard of competent writing of high quality, and more importantly as the way that native speakers commonly produce written products. Cortes (2004) identified that one of the most frequently occurring types of lexical bundles in L2 writing is lexical chunks composed of ‘noun phrases with a phrase fragment’ such as *the purpose of*, which is one of the habitual linguistic behaviors shared by English translators.

Table 2. Recurring lexical bundles in linguistics

Lexical Verb Phrase								
Rank	AE_LING	Freq.	Rank	EE_LING	Freq.	Rank	KE_LING	Freq.
1	it is argued	82	14	results showed that	47	4	this paper is	114
4	is argued that	69	17	results show that	43	7	paper is to	91
10	i argue that	41	21	the results showed	39	9	this study is	76
12	it is shown	40	22	show that the	37	10	study is to	71
14	is shown that	37	26	this study investigates	36	19	is to investigate	81
17	we argue that	35	28	study is to	35	32	this paper aims	37
18	argue that the	34	29	this paper is	35	34	is to examine	36
30	argues that the	30	33	this study is	34	48	paper aims to	33
43	this article argues	30	41	is to investigate	31			
44	argued that the	30						
49	article discusses the	31						
Lexical Noun-of Phrase								
Rank	AE_LING	Freq.	Rank	EE_LING	Freq.	Rank	KE_LING	Freq.
5	the use of	67	5	the effect of	59	1	the purpose of	161
6	the development of	56	6	the purpose of	59	2	purpose of this	154
8	analysis of the	45	9	the use of	57	11	the results of	70
11	the role of	41	11	purpose of this	53	18	analysis of the	52
16	the history of	36	15	the results of	47	23	the effects of	48
20	the basis of	34	23	different types of	36	24	the use of	48
			37	analysis of the	32	30	the effect of	38
			39	results of the	32	31	the analysis of	37
			47	the semantics of	30	36	results of the	36
						38	an analysis of	35
						40	the development of	35
						41	the basis of	33
						45	the frequency of	32
						49	the meaning of	30

4.3 Universals of syntactic explication: connectives

The numerical values of connectives in both disciplines came out in sequential order. Overall, the use of connectives seemed to be relatively standardized based on the observation of the raw data values. Across the CCERA, the KE_LIT corpus possessed the most, but it was not higher than what the KE_LING contained. The further investigation demonstrated that the mean values of each subcorpus in linguistics was statistically significant ($\chi^2(2)=71.872$, $p<.001$). The group differences in linguistics also reported statistical significance as the nontranslated corpus possess a smaller portion of connectives than that of both the quasi-translated and translated corpora (AE-EE: $p<.001$, AE-KE: $p<.001$, EE-KE: $p<.001$). In English literature, the mean values were statistically significant in all the three subcorpora ($\chi^2(2)=37.200$, $p<.001$). Likewise, paired corpora sets were also found to be statistically significant in their mean differences (AE-EE: $p<.001$, AE-KE: $p<.001$, EE-KE: $p=.030$). Overall, both linguistics and English literature were found to have a higher portion of connectives in translated texts than their equivalents, meaning that the observation was compatible with the hypothesis of explication. Further analyzed qualitatively, the arrangement of the discrete items covered in the list of the top-10 connectives paid particular attention (see Table 3). Then the regular and irregular patterns of the frequently encountered connectives were compared manually.

Table 3. The composition of connectives across the CCERA

The Top-10 Connectives in Linguistics											
Rank	AE_LING	Freq.	per 1,000	Rank	EE_LING	Freq.	per 1,000	Rank	KE_LING	Freq.	per 1,000
1	however	98	1.61	1	however	135	1.74	1	however	147	1.82
2	as well as	82	1.24	2	as well as	86	1.11	2	first	80	0.99
3	in addition	35	0.53	3	in addition	52	0.67	3	second	71	0.88
4	finally	32	0.49	4	finally	46	0.59	4	as well as	63	0.78
5	moreover	25	0.38	5	first	45	0.58	5	therefore	49	0.61
6	specifically	23	0.35	6	second	40	0.52	6	third	48	0.59
7	in particular	21	0.32	7	furthermore	35	0.45	7	furthermore	43	0.53
8	furthermore	15	0.23	8	specifically	33	0.43	8	finally	30	0.37
9	second	15	0.23	9	as a result	30	0.39	9	on the other hand	30	0.37
10	on the other hand	14	0.21	10	on the other hand	28	0.36	10	in addition	29	0.36
	⋮				⋮				⋮		
	TOTAL	605	5.77		TOTAL	870	8.19		TOTAL	1,065	10.00

The Top-10 Connectives in English Literature											
Rank	AE_LIT	Freq.	per 1,000	Rank	EE_LIT	Freq.	per 1,000	Rank	KE_LIT	Freq.	per 1,000
1	however	91	1.34	1	however	150	1.79	1	however	181	2.05
2	as well as	76	1.12	2	as well as	65	0.78	2	as well as	88	1.00
3	but	38	0.56	3	but	37	0.44	3	but	61	0.69
4	and then	21	0.31	4	thus	34	0.41	4	therefore	45	0.51
5	thus	20	0.30	5	on the other hand	30	0.36	5	in addition	31	0.35
6	yet	19	0.28	6	first	25	0.30	6	on the other hand	25	0.28
7	for example	18	0.27	7	in conclusion	24	0.29	7	despite	25	0.28
8	if not	17	0.25	8	in addition	24	0.29	8	so	25	0.28
9	moreover	16	0.24	9	meanwhile	21	0.25	9	as a result	22	0.25
10	in addition	16	0.24	10	therefore	19	0.23	10	accordingly	20	0.23
	⋮				⋮				⋮		
	TOTAL	585	5.53		TOTAL	806	7.61		TOTAL	877	8.44

In linguistics, the subcorpora had a quite unified pattern of recurrent connectives up to the first half of the list. However, the rest of the composition showed a somewhat heterogeneous preference, thus manifestly indicating the discrepancy in the lexical choices made by writers in each subcorpus. The traits evident in linguistics seemed to be no different from those of the other discipline. In English literature, the composition of high-frequency connectives ranked up to the third place was identical to one another, whereas the rest seemed to be quite manifold. To be specific, Korean scholars' texts showed a natural preference for *however* over the other connectives, which was not discernible in the nontranslated corpora in both disciplines. In Korean writers' subcorpora in English literature, the frequency of *however* all exceeded over 150 occurrences, of which the KE_LIT corpus seemed to be the most preoccupied with the use of the conjunction, having no alternatives at all. By comparison, in the nontranslated corpus, *however* was superseded by *yet* as ranked sixth or *if not* as ranked seventh instead. The similar trend was also noticeable in linguistics, having the frequency of 147 in the KE_LING and 135 in the EE_LING corpora.

4.4 Universals of syntactic convergence: mean sentence length SDs

During the stage of the baseline descriptive statistics, the mean sentence length of each subcorpus was counted to estimate the syntactic complexity of

each subcorpus. In linguistics, the nontranslated subcorpus had the highest value whereas the translated group had the lowest (AE_LING: 27.45, EE_LING: 26.02, KE_LING: 23.69). A similar pattern was also detectable in English literature (AE_LIT: 28.93, EE_LIT: 26.52, KE_LIT: 24.66). Further statistical analyses confirmed to the universals of syntactic simplification in linguistics ($\chi^2(2)=99.743$, $p<.001$), as well as in English literature ($\chi^2(2)=82.307$, $p<.001$). On the similar continuum of syntactic simplification, using a parameter of convergence, the standard deviations of mean sentence length were computed. In linguistics, the mean values of each subcorpus showed statistically significant results ($\chi^2(2)=72.718$, $p<.001$). Post-hoc tests proved that mean differences were statistically significant in all the paired datasets as well (AE-EE: $p<.001$, AE-KE: $p<.001$, EE-KE: $p<.001$). The English literature corpora also went through the same analyzing procedures as the other domain. Results reported that mean values were statistically significant ($\chi^2(2)=78.279$, $p<.001$). Mean differences were also found to be statistically significant (AE-EE: $p<.001$, AE-KE: $p<.001$, EE-KE: $p=.032$). As shown, there seem to be plausible results drawn as a sign of translation universals. Both disciplines were found to be consistent with the convergence hypothesis. The translated corpora of both fields appeared to hold lower values of standard deviations of mean sentence length than their counterparts.

All in all, most parameters were well indicative of lexical simplification, lexical normalization, syntactic explication, and syntactic convergence. The following matrix summarizes the results of the statistical analyses.

Table 4. Summary of statistical analysis results

LINGUISTICS				KRUSKAL WALLIS		POST-HOC			TRANSLATION UNIVERSALS	
VARIABLE	CORPUS	M	SD	χ^2	P	AE-EE	AE-KE	EE-KE	TESTING	CCERA DATASET
STTR	AE_LING	90.52	2.84	96.008	$p<.001$	$p<.001$	$p<.001$	$p<.001$	Hypothesis	NT > QT NT > TT QT > TT
Lexical Simplification (Lexical Variety)	EE_LING	89.49	2.97						CCERA DataSet	AE > EE AE > KE EE > KE
	KE_LING	88.68	3.06						Verification	Supported Supported Supported
Bottom-Frequency	AE_LING	39.64	0.08	146.970	$p<.001$	$p<.001$	$p<.001$	$p<.032$	Hypothesis	NT > QT NT > TT QT > TT
Lexical Simplification: (Lexical Sophistication)	EE_LING	35.31	2.97						CCERA DataSet	AE > EE AE > KE EE > KE
	KE_LING	34.15	3.06						Verification	Supported Supported Supported
Top-Frequency	AE_LING	51.42	0.05	142.730	$p<.001$	$p<.001$	$p<.001$	$p<.001$	Hypothesis	NT < QT NT < TT QT < TT
Lexical Simplification: (Lexical Richness)	EE_LING	53.17	0.05						CCERA DataSet	AE < EE AE < KE EE < KE
	KE_LING	55.38	0.06						Verification	Supported Supported Supported
Lexical Bundles	AE_LING	35.01	0.11	99.797	$p<.001$	$p<.001$	$p<.001$	$p<.001$	Hypothesis	NT < QT NT < TT QT < TT
Lexical Normalization	EE_LING	36.22	0.10						CCERA DataSet	AE < EE AE < KE EE < KE
	KE_LING	37.65	0.10						Verification	Supported Supported Supported
Connectives	AE_LING	0.94	0.01	71.872	$p<.001$	$p<.001$	$p<.001$	$p<.001$	Hypothesis	NT < QT NT < TT QT < TT
Syntactic Explication	EE_LING	1.10	0.01						CCERA DataSet	AE < EE AE < KE EE < KE
	KE_LING	1.35	0.01						Verification	Supported Supported Supported
Mean Sentence Length SDs	AE_LING	13.24	5.80	72.718	$p<.001$	$p<.001$	$p<.001$	$p<.001$	Hypothesis	NT > QT NT > TT QT > TT
Syntactic Convergence	EE_LING	11.73	4.78						CCERA DataSet	AE > EE AE > KE EE > KE
	KE_LING	11.62	6.86						Verification	Supported Supported Supported

ENGLISH LITERATURE				KRUSKAL WALLIS		POST-HOC			TRANSLATION UNIVERSALS			
VARIABLE	CORPUS	M	SD	X ²	P	AE-EE	AE-KE	EE-KE	TESTING	CCERA DATASET		
STTR	AE_LIT	90.91	2.55	36.999	<i>p</i> <.001	<i>p</i> <.001	<i>p</i> <.001	<i>p</i> =.150	Hypothesis	NT > QT	NT > TT	QT > TT
Lexical Simplification	EE_LIT	90.25	2.38						CCERA Dataset	AE > EE	AE > KE	EE > KE
Lexical Variety	KE_LIT	89.90	2.66						Verification	Supported	Supported	Rejected
Bottom-Frequency	AE_LIT	43.43	0.01	112.930	<i>p</i> <.001	<i>p</i> <.001	<i>p</i> <.001	<i>p</i> =.460	Hypothesis	NT > QT	NT > TT	QT > TT
Lexical Simplification:	EE_LIT	39.03	0.02						CCERA Dataset	AE > EE	AE > KE	EE < KE
Lexical Sophistication	KE_LIT	39.11	0.02						Verification	Supported	Supported	Supported
Top-Frequency	AE_LIT	46.95	0.05	19.194	<i>p</i> <.001	<i>p</i> <.001	<i>p</i> <.001	<i>p</i> =.750	Hypothesis	NT < QT	NT < TT	QT < TT
Lexical Simplification:	EE_LIT	48.03	0.04						CCERA Dataset	AE < EE	AE < KE	EE < KE
Lexical Richness	KE_LIT	48.05	0.05						Verification	Supported	Supported	Rejected
Lexical Bundles	AE_LIT	30.12	0.10	5.149	<i>p</i> =.076	<i>p</i> =.140	<i>p</i> =.140	<i>p</i> =.780	Hypothesis	NT < QT	NT < TT	QT < TT
Lexical Normalization	EE_LIT	26.66	0.08						CCERA Dataset	AE > EE	AE > KE	EE > KE
	KE_LIT	26.62	0.08						Verification	Supported	Supported	Supported
Connectives	AE_LIT	1.02	0.01	37.200	<i>p</i> <.001	<i>p</i> <.001	<i>p</i> <.001	<i>p</i> =.030	Hypothesis	NT < QT	NT < TT	QT < TT
Syntactic Explication	EE_LIT	1.06	0.01						CCERA Dataset	AE < EE	AE < KE	EE < KE
	KE_LIT	1.13	0.01						Verification	Supported	Supported	Supported
Mean Sentence Length	AE_LIT	14.63	6.24	78.279	<i>p</i> <.001	<i>p</i> <.001	<i>p</i> <.001	<i>p</i> =.032	Hypothesis	NT > QT	NT > TT	QT > TT
SDs	EE_LIT	12.96	4.65						CCERA Dataset	AE > EE	AE > KE	EE > KE
Syntactic Convergence	KE_LIT	12.24	4.95						Verification	Supported	Supported	Supported

The numerical figure in parentheses indicates the sample size of each group: AE_LING (N=600), EE_LING (N=605), KE_LING (N=603), AE_LIT (N=530), EE_LIT (N=440), and KE_LIT (N=435).

5. Discussion

Perceiving that L2 writers' texts may bear neither resemblance with those of nontranslated comparable counterparts nor relation to the essential attributes of translation, this article argued that the distinctive properties of translational English may delineate the hallmarks of L2 writing produced by non-anglophone scholars in the disciplines of linguistics and English literature. Hinged on the hypotheses of translation universals postulated by Baker (1996), the present study addressed how salient translational features arose in the expert L2 writers' texts generated by Korean scholars. The following summarizes some of the critical investigations by pinpointing subcorpus-specific results along with discipline-specific findings.

On a substantial level, a general presumption on the interrelatedness between Korean scholars' L2 writing and translation universals turned out to be warranted. Setting plausible parameters to discern translational instances related to lexical and syntactical choices, this article argued that the hallmarks of Korean scholars' L2 writing may comprise the translational features of lexical simplification, lexical normalization, syntactic explication, and syntactic convergence in their broad outlines. As expected, Korean scholars' English abstracts were lexically simplified and normalized, as well as syntactically explicated and converged to the fullest extent when compared to native scholars' journal abstracts.

The first sub-hypothesis regards the lexical simplification in the universals of translation, and three indices were tested to trace whether Korean L2 writers' texts are lexically simplified or not. One indicator used to test the universals of lexical simplification was the STTR values so that lexical variety could evaluate how various words run across the CCERA. As Baker (1996) affirmed, translated and quasi-translated L2 writings expectedly had smaller values than nontranslated English texts, which therefore caused the lexical diversity of Korean scholars' L2 texts to fall lower (e.g., Laviosa 2002). Korean scholars' texts showed evidence of lower lexical diversity, coinciding with the premise that L2 writers might have recycled highly recurring words repeatedly to meet the same length of texts as their counterparts, thus being lexically less diverse. Also, the AE_LIT corpus represented the greatest lexical diversity across the CCERA. The pair of the EE_LIT and KE_LIT corpora was the only one that showed an insignificant relationship. As for discipline-specific findings, the lexical diversity in the linguistics domain was relatively lower than that in the English literature corpora. It seems plausible that the translational parameter of STTR has well accounted for the tendency of lexical simplification as one of the translational manifestation salient in Korean scholars' writings. All around, these results associated with the fact that when manipulating their mental lexicon, Korean L2 writers in the discipline of linguistics seems to be less dauntless compared to the scholars in the other discipline. Another indicator applied for lexical simplification is related to the concept of lexical sophistication. Consistent observations employing the bottom-frequency values were discernable as a sign of translation universals of simplification. The subcorpora in both domains were consistent with the hypothesis of translation universals. The group differences between nontranslated and quasi-translated/translated corpora were quite evident so that the parameter of bottom-frequency words proved to be effective, which in turn corroborates the premise that Korean scholars' L2 texts may have a lower level of lexical sophistication than native scholars' English texts. The last indicator of lexical simplification is pertinent to lexical richness (Baker 1996). The portion of top-frequency words as a measurement variable demonstrated that translated and quasi-translated L2 texts contained smaller numbers of top-frequency words than their counterparts. In general, using the measurement variable of STTR, translated L2 texts appeared to have the lowest level of lexical

richness, reaffirmed by the analysis of lexical variety. In particular, translated L2 texts in linguistics marked the highest amount of recurring words, and quasi-translated texts recorded the second highest across the CCERA; thus, their lexis being comparatively least and less well-off. The similar case was recognizable in English literature but not significant between the quasi-translated and translated corpora. For example, such shifts were prevalent in the use of nouns and verbs when stating the purpose of abstract writing. Translated texts in both disciplines adopted a limited and repetitive manner in the use of nouns and verbs for a research aim, showing an inclination to nouns rather than verbs.

The second sub-hypothesis regards the values of lexical bundles in compliance with the universals of lexical normalization (Baker 2007; Olohan 2004). Korean writers' L2 texts had a higher portion of recurrent lexical bundles than their counterparts in both disciplines. Concurrently, the quasi-translated texts in linguistics showed robust evidence regarding recurrent lexical bundles. Extra attention was also paid to the overall frequency of verb phrases. Results indicated that nontranslated corpora had the most varied types of verb phrases and the smallest types of *noun-of* phrases while translated linguistics corpora contained the smallest types of verb phrases and the greatest types of *noun-of* phrases. Results indicated that translated L2 writings preferred to use more of prefabricated collocations, thereby proving that Korean scholars' L2 texts bear translational normalization (e.g., Øveras 1998).

The third sub-hypothesis of syntactic explicitation was traced using the proportion of connectives running across the CCERA. As the hypothesis of explicitation assumes (Blum-Kulka 1986; Øveras 1998; Olohan and Baker 2000), translated texts produced by expert L2 writers held conjunction words more than their counterparts (e.g., Xiao and Dai 2014). Amongst the six subcorpora, translated linguistics corpus contained the highest rate of connectives, translated English literature had the second largest, and quasi-translated linguistics was ranked at the third place from the top. Besides the frequency of connectives occurring in each corpus, the compositional differences of the top ten high-frequency connectives were also examined. The use of connectives seemed to be quite arbitrary among translated and nontranslated L2 texts, except for the use of *however* with discernable preference apparent in Korean scholars' L2 texts. Whatever connectives were preferred, Korean L2 writers' texts appeared to

maintain the same linguistic features as those of typical translated texts accordingly. Having said this, these findings might be controversial in terms of three different aspects. On the one hand, the current finding is in line with the perspective of the linguistic features of L2 writers' texts because previous L2 studies have suggested that more skilled L2 writers used more cohesive devices (e.g., Connor 1999). On the other hand, the L1 writing approach has encouraged writers to use fewer cohesive devices (McNamara et al. 2010, 2011). In the view of rhetorical analysis, these results could be quite implausible as well. Previous findings have postulated that competent writers, either L1 or L2, tend to make cohesive links between ideas and clauses by utilizing connectives (Longo 1994). Undoubtedly, using connectives as a cohesive device seems to be instrumental for L2 writers to develop ideas and to reach a level of cohesion, clarity, and even maturity in their writing that is as high as that of native writers. In light of producing authentic texts with no features of translationese, however, one more additional premise needs to be considered in the future research attempts. That is, employing varied cohesive lexical ties in a 'conservative' way should be one of the primary features in nontranslated English texts, making a supposition that anglophone countries have high-context cultures (Hall 1976) where implicit communication is favored, so people benefit from less cohesive texts. Conversely, non-anglophone countries such as Korea have low-context cultures where explicit interaction and more cohesive texts are believed to be effective (O'Reilly and McNamara 2007). L2 abstract writing is more than just choosing the right words. Accordingly, it should ensure that the target audience in the written academic context receives the right messages and thoughts in the way the writers intended. L2 writing *per se* is still a long way from achieving authenticity; be that as it may, L2 writing does not always have to share the same linguistic features inherent to most conventional translated texts, in any case.

The last sub-hypothesis of syntactic convergence as translation universals was also evaluated using the standard deviations of mean sentence length (Laviosa 2002). The Korean L2 writers in the linguistics corpora showed a lower standard deviation of mean sentence length than their counterparts (e.g., Yajun and Zaixin 2008). Korean L2 writers' linguistic proximity could be supported by the predictor, resulting in the higher level of homogeneity among the same corpus group. Initially proved by the universals of simplification, Korean L2 writers'

corpora showed quite a similar level of writing proficiency with shorter length of sentences than nontranslated texts. As the results lean towards what the convergence hypothesis posits, Korean scholars' L2 texts had a distinctive array of linguistic closeness across the subcorpora. Specifically, judging from the fact that a standard deviation quantifies the amount of 'variation or dispersion' of the given values of syntactic difficulties, Korean L2 writings in the English Literature discipline seemed to be somewhat balanced out across the corpora. The values of the three groups even showed a similar level to that of the nontranslated linguistic corpus. As observed, the linguistic proximity between Korean L2 writers in the English literature domain was confirmed by the predictor, therefore leading to the higher level of homogeneity than that of their equivalents. It seems reasonable to assume that the Korean L2 writers' texts in these corpora may well be evened out in the aspect of syntactic variation and dispersion. Conformed to the expectation, the standard deviations of mean sentence length in both Korean scholars' abstracts corroborated the translation universals of convergence. As uncovered by the parameter of convergence, the universal features of translation seem to be pervasive, to a larger extent, in Korean L2 writers' texts.

Concerning the two research questions on linguistic features and indices associated with translational universals that make Korean L2 writings distinctive to native speakers' original writing, the following section underscores the four major qualities of L2 writers' texts along with valid translational indices unique to Korean L2 writing. First, the tendency of lexical simplification was found to be universal, rather than norm-dependent. Korean L2 writers' texts contained far much lower lexical variety, lower lexical sophistication, and lower lexical richness than native speakers' nontranslated original texts so that the three indices were found to be valid, except for the corpus pair of quasi-translated and translated texts. The second linguistic feature shared in L2 writers' texts would be the 'propensity' of using specific word partners, namely the excessive use of *noun-of* bundles appeared to be another strong 'universal' property apparent in Korean L2 writers' texts. The use of these lexical bundles in a mechanical way seemed to be more manifest, signifying the nature of Korean L2 writers' texts as the product of translational event driven by the cognitive process of L2 writing. Third, as the previous studies strongly posited, the universals of explicitness in

L2 writers' texts were quite convincing. The only unexpectedness would be the strong counter-evidence to the proposition that L2 writers' texts at high proficiency levels are less likely to be explicit so that these texts might contain fewer connectives, which was not evident in this study. Even though the L2 writings are those by 'expert' L2 writers, their writing still seems to bear similar linguistic traits inherent to translational English in that they have a high proportion of connectives (e.g., Connor 1999; Hinkel 2002; Swales 1990; Qi 1986). As a consequence, the orientation towards explicitness regarding the use of connectives seems a convincing instance perceptible to expert L2 writers' texts.

By and large, the instances of translational manifestations detected in Korean L2 writing can be explained to be valid following linguistic behavior and L2 language processing. Greater awareness towards L2 writing product driven by L2 writing process may promote expert L2 writers to become more consciously aware of their lexical choices and language behaviors that would contribute to the quality of their L2 writing accompanied by a mental translation. As Scott (1996) claims, L2 writers are expected to make all the necessary linguistic choices incorporated with lexical items, syntactic structures, and stylistic constructions to transform their thoughts into the target language. Dominated by the cognitive function of L2 writers, the restricted mode of mental translation seems to be no other than an 'inescapable cognitive mechanism' closely intermingled with translation as a conscious or unconscious cognitive process (Chesterman 2004b, 2006). Seeing that no matter what language backgrounds they have and no matter how their writing proficiencies are described, every L2 writer seems to be destined to go through the cognitive process of L2 writing driven by mental translation (e.g., Cook 1992). Even if Korean scholars' English abstracts were not perfect translations *per se*, their journal articles seemed to perform a role of a source text as if L2 writing were conventional regular translating work. Consequently, mental translation every L2 writer had to pass through seemed to give rise to 'unwilled translationese' inevitably marked on their writings.

All told, the findings of this study seem to coincide with the notion of translational manifestations. Concurrently, the translational manifestations, which were quite pervading in expert L2 writers' texts, might be emanated from bidirectional links: mental translation as both the 'process' and the 'product' of second language writing. It is to this end that the following section accounts for

whether an act of 'translation' taken during the cognitive 'process' of L2 writing affects the 'product' of L2 writing ensued, by which involves translational language manifested in expert L2 writers' texts. In the first light of mental translation as a process, it is plausible to interpret that the process of a mental translation could have affected the 'product' of L2 writing by infiltrating translationese, especially when making lexical choices. Although Wang and Wen's (2002) L2 writing model describes wherein the composing processor and what language dominates the process of L2 writing, it seems still insufficient in that it does not touch the heart of the matter, 'downside of translation.' That is, their model only accounted for the positive role of translation as a catalyst to accomplish the 'act of L2 writing' *per se*. It is noteworthy that 'translation-like' lexical and syntactic choices were also drawn by the "consequences of either a deliberate strategy of translation or due to largely unconscious cognitive processing that forms part of the complex nature" (Olohan 2001: 423).

The second link highlights mental translation as a product in L2 writing. It seems probable that Korean scholars' L2 texts driven by the process of L2 writing could be a large part of the 'by-products of mental translation.' It is simply attributable to the fact that a significant number of signals representing translational manifestations were unquestionably noticeable in expert L2 writers' texts. Even though the L2 writers' texts are not the outcomes produced during the conventional process of translation, those texts are no other than the final 'products' arisen by mental translation. In this respect, it seems feasible to accept the translational manifestations of L2 writing as sort of a 'spin-off' provoked during the L2 writing process (e.g., Baker 1993; Lee 2017), which would thus represent linguistic attributes pertinent to L2 writers' texts.

6. Conclusion and implications

The findings of the study have diverged from the conventional understandings concerning the linguistic qualities of L2 writing in the Korean context. This article, therefore, claims that linguistic features of expert L2 writers' texts are closely linked to translational strains but contrary to those of native speakers' written production, thereby disapproving a commonly held notion that

skilled L2 writers would produce L2 texts which rarely bear the common features of translational English. By raising the awareness of translational properties apparent in L2 writers' texts, expert L2 writers will have a deeper understanding on the distinct nature of L2 writing which is affected by mental translation as a process and as a product. Going one step further, by facilitating better understanding on what linguistic features distinguish expert L2 writers' texts from native writers' original texts, those who manipulate the English language as their L2 may broaden their working knowledge repertoires. While elaborating their L2 texts in practical context, expert L2 writers can challenge themselves to reach up to the level of linguistic authenticity and 'textual fit' (e.g., Chesterman 2004a). The final pedagogical aim building metacognitive awareness to be mindful of their L2 language processing, thus, expert L2 writers need to equip themselves with metacognitive strategies so that they can be consciously and explicitly aware of what to avoid and what to accept during the L2 writing process entailing mental translation. The metacognitive knowledge built upon the solid evidence confirmed through the present study would enable Korean L2 writers as 'cognitive processors' to become more efficient and, importantly, more autonomous as an English language learner and user (see Lee 2017).

Given that mental translation may affect the product of L2 writing, consciously or unconsciously, it is worthwhile to note that L2 writers in non-anglophone cultures may need to concede that L2 writing is never detachable from translation, causing translational properties to be shared by translators' texts. More importantly, the truth that these sort of L2 texts may bear the features of translationese should never be considered something disgraceful. Instead, such texts should be 'dignified as a variant understanding of second language writing.' One apparent reason for this is that, as noted earlier, second language writing triggered by a mode of mental translation will broaden our knowledge on L2 writers' texts as its subfield, and by that means, will add another diversity to different varieties of World Englishes.

Despite some limitations and deficiencies, this study may construct a plausible answer to the query addressed by Chesterman (2004b) regarding the notion of translation universals in that translational features may be not only pertinent to translating context but also apply to extra-constrained communication contexts such as 'L2 writing' or 'L2 speaking.' In an effort to

augment the validity of the findings, it is hoped to implement further research based on the interdisciplinary studies combining empirical approaches using corpora and psycholinguistic investigation, in addition to an embodiment of biological data within the foundation of cognitive science. Such research endeavors will undoubtedly have immense potential to contribute to the understanding of the unknown relationship between L2 writing and translation as we do not know yet precisely to what extent the occurrence of translational manifestations in L2 writers' texts is only parts of the bigger picture.

References

- Anthony, Lawrence. 2013. A critical look at software tools in corpus linguistics. *Linguistic Research* 30(2): 141-161.
- Bagheri, Mohammad Sadegh and Ismaeil Fazel. 2011. EFL learners' beliefs about translation and its use as a strategy in writing. *Reading Matrix: An International Online Journal* 11(3): 292-301.
- Baker, Mona. 1993. Corpus linguistics and translation studies: Implications and applications. In Mona Baker, Gill Francis and Elena Tognini-Bonelli (eds.), *Text and technology: In honour of John Sinclair*, 233-250. Amsterdam: John Benjamins.
- Baker, Mona. 1995. Corpora in translation studies: An overview and some suggestions for future research. *Target* 7(2): 223-243.
- Baker, Mona. 1996. Corpus-based translation studies: The challenges that lie ahead. In Harold Somers (ed.), *Terminology, LSP and translation*, 175-186. Amsterdam and Philadelphia: John Benjamins.
- Baker, Mona. 2004. A corpus-based view of similarity and difference in translation. *International Journal of Corpus Linguistics* 9(2): 167-193.
- Baker, Mona. 2007. Patterns of idiomaticity in translated vs. non-translated text. *Belgian Journal of Linguistics* 21(1): 11-21.
- Biber, Douglas and Susan Conrad. 1999. Lexical bundles in conversation and academic prose. *Language and Computers* 26: 181-190.
- Blum-Kulka, Shoshana. 1986. Shifts of cohesion and coherence in translation. In Juliane House and Shoshana Blum-Kulka (eds.), *Interlingual and intercultural communication: discourse and cognition in translation and second language acquisition studies*, 17-35. Tübingen: Gunter Narr.
- Chae, Soo Eun. 2012. Identifying effective English L2 writing interventions: Emerging trends and issues in recent research. *English Teaching* 67(4): 3-24.

- Chesterman, Andrew. 2006. Interpreting the meaning of translation. In Mickael Suominen (ed.), *A man of measure festschrift in honour of Fred Karlsson on his 60th birthday*, 3-11. Turku: Linguistic Association of Finland.
- Chesterman, Andrew. 2004a. Beyond the particular. In Anna Mauranen and Pekka Kujamaki (eds), *Translation universals: Do they exist?*, 33-49. Amsterdam: Benjamins.
- Chesterman, Andrew. 2004b. Hypotheses about translation universals. In Gyde Hansen, Kirsten Malmkjar, and Daniel Gile. (eds), *Claims, changes and challenges in translation studies*, 1-13. Amsterdam: John Benjamins.
- Connor, Ulla. 1999. Linguistic/rhetorical measures for international persuasive student writing. *Research in the Teaching of English* 24(1): 67-87.
- Cook, Vivian. 1992. Evidence for multicompetence. *Language Learning* 42(4): 557-591.
- Crossley, Scott Andrew and Danielle McNamara. 2010. Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading* 35(2): 1-21.
- Crossley, Scott Andrew and Danielle McNamara. 2011. Shared features of L2 writing: Intergroup homogeneity and text classification. *Journal of Second Language Writing* 20(4): 271-285.
- Cumming, Alister. 1990. Metalinguistic and ideational thinking in second language composing. *Written Communication* 7(4): 482-511.
- Cumming, Brett. 2012. A critical analysis of writing competence and characteristic errors of Japanese EFL university majors: Implications for strategic teaching. *Aichi Prefectural University Mulberry* 61: 159-185.
- Flower, Linda and John Hayes. 1981. A cognitive process theory of writing. *College Composition and Communication* 32: 365-387.
- Goh, Gwangyoon and Younghee Cheri Lee. 2016. A corpus-based study of translation universals in English translations of Korean newspaper texts. *Cross-Cultural Studies* 45: 109-143.
- Goh, Gwangyoon, Younghee Cheri Lee, and Dongyoung Kim. 2016. A Corpus-based study of translation universals in thesis/dissertation abstracts. *Korean Journal of English Language and Linguistics* 16(4): 819-849.
- Grabe, William. 2001. Notes toward a theory of second language writing. In Tony Silva and Paul Matsuda (eds.), *On second language writing* 39-57. Mahwah, NJ: Erlbaum.
- Grabe, William and Robert Kaplan. 1996. Theory and practice of writing: An applied linguistics perspective. Harlow: Pearson Education.
- Gravetter, Federiock and Larry Wallnau. 2013. *Statistics for behavioral sciences*. 9th Edition. Belmont, CA: Wadsworth.
- Hall, Edward. 1976. *Beyond culture*. Garden City, NY: Anchor Press.
- Hinkel, Eli. 2002. *Second language writers' text: Linguistic and rhetorical features*. Routledge. Lawrence Erlbaum Associates.

- Hyland, Ken. 2002. Activity and evaluation: Reporting practices in academic writing. In John Flowerdew (ed.), *Academic Discourse*, 115-130. London: Longman.
- Kaplan, Robert Barbara. 2000. Response to "On the future of second language writing", *Journal of Second Language Writing* 9(3): 311-314.
- Kellogg, Ronald. 1987. Effects of topic knowledge on the allocation of processing time and cognitive effort to writing processes. *Memory and Cognition* 15(3): 256-266.
- Kim, Hyoung-il. 2010. A study of the frequency of linguistic features in Korean students' L2 writing. *English Language Education* 40: 110-126.
- Laviosa, Sara. 1998a. The corpus-based approach: A new paradigm in translation studies. *Meta: Translators' Journal* 43(4): 474-479.
- Laviosa, Sara. 1998b. Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta: Translators' Journal* 43(4): 557-570.
- Laviosa, Sara. 2002. *Corpus-based translation studies: Theory, findings, applications*, Vol. 17. New York: Rodopi.
- Lee, Eunpyo. 2001. A study on error analysis and hedging expressions of medical research abstracts. *English Language and Literature Teaching* 13(1): 47-66.
- Lee, Younghee Cheri. 2017. The hallmarks of expert L2 writers' texts viewed through the prism of translation universals: A corpus-based approach to English research abstracts of scholarly journal articles. PhD Dissertation. Yonsei University.
- Leki, Ilona, Alister Cumming, and Tony Silva. 2010. *A synthesis of research on second language writing in English*. New York: Routledge.
- Longo, Bernadette. 1994. The role of metadiscourse in persuasion. *Technical Communication* 41(2): 348-352.
- Oi, Kyoko. 1986. Cross-cultural differences in rhetorical patterning: A study of Japanese and English. *The Japan Association of College English Teachers*, (17): 23-48.
- Olohan, Maeve. 2001. Spelling out the optionals in translation: A corpus study. *UCREL Technical Papers*, 13: 423-432. Lancaster: UCREL, Lancaster University.
- Olohan, Maeve. 2004. *Introducing corpora in translation studies*. London: Routledge.
- Olohan, Maeve and Mona Baker. 2000. Reporting 'That' in translated English: evidence of or subliminal processes of explicitation. *Across Languages and Cultures* 1(2): 141-158.
- O'Reilly, Tenaha and Danielle McNamara. 2007. Reversing the reverse cohesion effect: good texts can be better for strategic, high-knowledge readers. *Discourse Processes* 43(2): 121-152.
- Øverås, Linn. 1998. In search of the third code: An investigation of norms in literary translation, *Meta: Translators' Journal* 43(4): 571-588.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Sasaki, Miyuki. 2000. Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing* 9(3): 259-291.

- Schoonen, Rob, Amos Van Gelderen, Kees De Glopper, Jan Hulstijn, Annegien Simis, Patrick Snellings, and Marie Stevenson. 2003. First language and second language writing: The role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language Learning* 53(1): 165-202.
- Scott, Mike. 2017. WordSmith tools version 7, Stroud: Lexical Analysis Software.
- Scott, Virginia Mitchell. 1996. *Rethinking foreign language writing*. Boston, MA: Heinle and Heinle.
- Silva, Tony and Colleen Brice. 2004. Research in teaching writing. *Annual Review of Applied Linguistics* 24: 70-106.
- Swales, John. 1990. *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Uzawa, Kozue. 1996. Second language learners' processes of L1 writing, L2 writing, and translation from L1 into L2. *Journal of Second Language Writing* 5(3): 271-294.
- Wang, Wenyu and Qiufang Wen. 2002. L1 use in the L2 composing process: An exploratory study of 16 Chinese EFL writers. *Journal of Second Language Writing* 11: 225-246.
- Woodall, Billy. 2002. Language-switching: Using the first language while writing in a second language. *Journal of Second Language Writing* 11(1): 7-28.
- Xiao, Richard and Guangrong Dai. 2014. Lexical and grammatical properties of translational Chinese: Translation universal hypotheses reevaluated from the Chinese perspective. *Corpus Linguistics and Linguistic Theory* 10: 11-55.
- Yajun, Jiang and Ren Zaixin. 2008. Translated English and universals of translation. *English Today* 24(4): 26-33.

Younghee Cheri Lee

Adjunct Lecturer
English Education
Graduate Program in Cognitive Science
Yonsei University
50 Yonsei-ro, Seodaemun-gu, Seoul 120-749, Korea
E-mail: cheriberry@yonsei.ac.kr

Received: 2018. 04. 21.

Revised: 2018. 06. 04.

Accepted: 2018. 06. 04.