Korean Journal of English Language and Linguistics, **Vol 21, March 2021, pp. 261-281**
DOI: 10.15738/kjell.21..202103.261

KOREAN JOURNAL OF ENGLISH LANGUAGE AND LINGUISTICS

ISSN: 1598-1398 / e-ISSN 2586-7474
http://journal.kasell.or.kr

Function Words as Markers of Translationese: A Corpus-based Approach to Mental Translation in Second Language Writing

Younghee Cheri Lee (Chung-Ang University)

Received: February 27, 2021
Revised: March 15, 2021
Accepted: March 25, 2021

Younghee Cheri Lee
Lecturer, Dept. of English Education, Chung-Ang University
Tel: 02-820-5391
cheriberry@cau.ac.kr

ABSTRACT

Lee, Younghee Cheri. 2021. Function words as markers of translationese: A corpus-based approach to mental translation in second language writing. *Korean Journal of English Language and Linguistics* 21, 261-281.

Driven by the linguistic values of closed-class words, this article seeks to provide a multifaceted account of function words as markers of translationese, thereby aiming to reconceptualize the universal trait of translational manifestations found in non-translated L2 writing. To that end, using comparable monolingual English corpora from two different disciplines, this study implemented a two-fold analysis to compare a conventional analytical model (i.e., "all-token variable" of function words) with a modified approach (i.e., "subset variables" of function words). The "all-token" method has been one of the most unstable measures in the studies of translation universals (TU) and still lacks a coherent understanding of how specific function words should be attested in their predictive roles in translationese. As contrasted with conventional TU assumptions, it was evidenced that the "all-token" function words outperformed only in a single domain, a result that distanced from the universal traits of translationese. Instead, as one of the subset variables, auxiliary verbs demonstrated a higher predictive and universal power as a newly attested translationese marker. Thus, this article argues that the notion of translationese should be reframed as "universal" translationese and "domain-specific" translationese, respectively. The rationale lies in that the predictive roles of subset function words have been overshadowed by the inconsistent analytical method implemented in translation studies to date.

KEYWORDS

function words, translationese, Translation Universals (TU), mental translation, second language (L2) writing, non-nativeness

1. Introduction

Function words account for only about 1% of English vocabulary, but they take up approximately 60% of the tokens produced in daily discourses (Pennebaker 2011). Although function words surpass the number of content words created every day, it is incredibly challenging for us to recognize them by the other party in any speech settings. Regardless of a discourse type, spoken or written, function words are easy to pass through in most cases, thus being somewhat negligible and hard to perceive. Similarly, it is almost impossible to detect function words intuitively that occur most frequently or least frequently through texts unless manually counted. It is not common to pay attention to function words unless listeners or readers are linguists or grammarians. Of course, even with the intuition of a native speaker, the case is no different. In a linguistic context, function words are grammatical words that signify the structural connections between words and words while working like a thread that holds up sentences together and has ambiguous or little lexical meaning in themselves. Function words are the least noticeable of all parts of speech in the English language, so we might not be conscious of their value, either. Therefore, such secrecy and the covert nature of function words seem to render their status to be assigned to a closed class, thereby grabbing less scholarly attention of themselves. Thanks to enhancing text analysis techniques and tools, however, these function words have recently gained scholarly attention in linguistics-related disciplines and neighboring disciplines such as psychology and cognitive science.

In particular, in the scholarly field of corpus-based translation universals (TU), where the in-depth validation of linguistic factors is attested, substantial research has been carried out by adopting function words as an operational parameter to spot translationese[1]. Originated by its forerunner Gellerstam (1986), translationese is any linguistic difference found between non-translated and translated texts. Regardless of specific language pairs, certain linguistic features that commonly arise in non-native writers' texts are also understood as translationese, in which a different linguistic taste of foreignness is pervasive (see Lee 2019). Over the past decades, translationese from the angle of translation universals (TU) has been the center of translation studies' research topics. Due to the algorithmic enhancement in machine learning and deep learning, a growing number of studies, mostly in computer linguistics, have been paying particular attention to translationese in the aspects of translation quality and fluency using the TU variables (see Secara 2005, Baroni and Bernardini 2006, Gaspari and Bernardini 2008, Koppel and Ordan 2011, Kunilovskaya and Lapshinova-Koltunski 2019).

In the domestic research context, however, few studies have paid close attention to the predictive values of function words as markers of translationese found in second language writing (i.e., non-translated L2 texts), most of whose findings are still unstable and wandering (see Lee 2017, 2018, 2019). More importantly, many prior studies have failed to truly explain function words' distinctiveness as the predictors of translationese compared to the use of content words as TU variables. In the debate over function words as unstable markers of translationese, the heart of the matter resides in how function words are operated as TU variables cause conflicts among TU hypotheses by failing to include a more thorough and unified description. Therefore, this study attempted to look closely into the predictive power of function words by varying analytic methods of how those factors should be treated in an effort to fundamentally attest to translationese with a new angle in L2 writing and thus unveil the inherent nature of non-nativeness.

[1] The term translationese is referred to as linguistic "fingerprints" that are left on one language from the other language by the process of translation. Lee (2018, 2019) has further extended its original definition to mean the "fingerprints of non-nativeness" or "foreignness" that are markedly permeable in non-translated L2 writers' texts (c.f. Newmark 1991, Toury 1979).

2. Literature Review

2.1 Function Words Across Disciplines

In the linguistics discipline, an initiator opened his eyes to the undisclosed value of function words that could be seemingly insignificant. Triggered by his seminal work in 1984, Biber (1988) finally laid the groundwork for research that exposed variations in styles and writing genres based on understanding "function words' attributes." In extensive empirical research with 67 linguistic factors utilized, he explored linguistic aspects that emerge accompanied by discourse function as an overarching principle but not grammatical function. His inductive approach provided clumped word patterns as per their normal co-occurrences in the factor-based analysis to discern linguistic properties that would explain language variations in different genres and texts. In the study of cross-linguistic variations (Biber 1995), in particular, a set of lexico-grammatical identifiers proposed by Biber (1988) was implemented to unveil a "cross-linguistic universals" of register variation with remarkable commonalities and distinctiveness in four language pairs with generic differences (English, Nukulaelae Tuvaluan, Korean, and Somali). Such research endeavor has been extended across disciplines to detect "translationese" in that there can be textual distinctiveness between translated and non-translated texts. Biber's such influential initiation has affected neighboring disciplines, driven by the proposition that words are generally considered to be clustered by their grammatical functions (Biber 1995, 1998, Pennebaker and King 1999). Evert and Neumann (2017) deployed Biber's spectrum of lexico-grammatical variables to discern register variations between translated texts and non-translated texts and reported linguistic differences between these two.

The value of function words can also be discussed concerning lexical density. Lexical density is one of the linguistic complexity measures based on the counting units of grammatical words and lexical words in speech or texts. There are quite different types of proposed methods for the calculation of lexical density. Laufer and Nation's (1995) model is most meticulously structured, defining lexical density as the percentage of lexical words running across the text. They presented a formula for calculating lexical density by using the ratio comparison method between function words and content words (i.e., nouns, verbs, adjectives, and adverbs) in research on lexical richness in L2 written production. Their formula has been reproduced in numerous studies to date. According to these researchers, lexical words mainly provide meaning, and a text is deemed to be dense if it contains many lexical words compared to the total token of function and lexical words. They argue that the portion of function words running across a text affects the text's structural features. For example, the text with fewer function words may bring about more subordinate clauses, participle phrases, and ellipses (Laufer and Nation 1995). Laufer and Nation's arguments for the lexical density measurement have been probably deeply engraved with Biber's (1988) insights, who prepared its core decades ago.

In the corpus linguistics domain, Kennedy (1998) also discussed the importance of using function words as a dependent variable. He stated that even with different sizes or genres of the two corpora, no substantial differences were noticeable in the composition or frequency of function words and the combination of high-frequency function words (Kennedy 1988, Leech, Rayson, and Wilson 2001, McEnery and Wilson 2001). It can be further perceivable that any text or speech generated at the level of native speakers should have a similar composition and scale of function words. Supporting Kennedy's claim, Goh (2007) claimed that the distribution of high-frequency function words could be helpful as an index to gauge the level of authenticity in L2 texts. High-frequency function words ranked at the top may represent those texts' fundamental grammatical features, thus being a very suitable parameter for evaluating the language pairs' linguistic quality compared across the texts.

In the psychology-related disciplines, various research endeavors have been made to identify the connection between thought and language use. Until the end of the twentieth century, language processing has not been considered a significant research methodology for the discipline of psychology. Most psychological experiments

have thus been confined to response time as dependent variables to explore several facets of psychological properties or associated cognitive processes (Pennebaker and Graybeal 2001). Since the advent of a language processing tool tailored for psychiatric investigations and clinical inquiries, such new study scenes have attracted increasing attention on language variables such as function words. Backed up by a vast majority of research into function words, more pertinently, Pennebaker (2003) pinpointed one very insightful claim that our brains (or minds) are naturally built to slip over function words but immediately and automatically concentrate on content-related words. It is well-known that the human brain has a language center taking charge of language functions, in which Broca's and Wernicke's areas along with angular gyrus are assigned with the production and processing of languages. In particular, the processing of function words is connected to the Broca's area, which is a place for syntactic processing and style control, whereas Wernicke's area controls content words for language comprehension, though not located in one set area (Hartley, Sotto, and Pennebaker 2002, Pennebaker 2011). Storing images associated with words, sociocultural connections, memories, and emotional connections, Broca's area is always working as we process language very subconsciously. More pertinently, function words are marked by swift processing in Broca's region (see Pennebaker 2003, 2007, 2011, 2013). In a similar vein, Hartley (2002) also asserted that our minds could determine which function words to use "almost instantaneously." In this regard, Pennebaker outlines the four distinctive attributes of function words by implying why those function words are of great importance in psychological and cognitive aspects (Pennebaker 2011: 23):

(1) Function words are used at very high rates.
(2) Function words are short and hard to detect.
(3) Function words are processed in the brain differently from content words.
(4) Function words are extremely social.

Given that function words are hard to be controlled even by native writers, what must be noted here is the potential value of function words to uncover the inherent disparities between native writers and non-native writers in the significant respect that non-native writers would not be able to control their use of function words as much as they could do with content words. Both translating and L2 writing activities are intertwined with unconscious or subconscious cognitive processes (e.g., Cook 1992, Hartmann and Stork 1972, Perani, Stanislas and Grassi 1996, Wang and Wen 2002, Qi 1998, Ransdell and Barbier 2002, Uzawa 1996). The exploration of function words, which are considered a by-product of unconscious language processing (Baker 1996), would eventually detect translationese in L2 English writings.

The present research is intended to identify function words' predictive values as translationese signifying non-nativeness by assessing language use differences between native and non-native writers. Translationese is a linguistic phenomenon in the "cognitive process" of writing (Baker 1996). The manifestations of mental translation[2] might occur in the skilled non-native writers' L2 texts (see Lee 2017, 2018, 2019, Perani et al. 1996). As such, it is evident that there are restraints to describing the products of L2 writing, whether direct or mental, without a complete understanding of human psychological and cognitive processes. Pennebaker's intriguing view to function words will help us deepen our understanding of human writing's psychological and cognitive processes, thus enabling us to unlock the secrets to the hallmarks of non-nativeness[3] as well as translationese.

[2] According to the Google Ngram Viewer, up until the late 2000s, the term "mental text" (Ransdell and Barbier 2002) was more favored than "mental translation" (Kern 1994). The former is centered around the product of translating, whereas the latter weighs more on the process of writing (cf. Lee 2018). Hönig (1991) explained the "mental translation" processes as an image led by the mental representation of a target text during the process of translation.

[3] The term non-nativeness is collectively referred as any linguistic properties markedly observable in non-translated L2 writings and perceptively different from those of L1 writers' original texts (Lee 2018, 2019).

2.2 Function Words as a Prism of Translationese

Studies on the different use of function words between native and non-native writers have been actively attempted to date, especially in descriptive translation studies. Corpus-based contrastive studies of translationese have centered mainly on the hypotheses of translation universals thus far. In the early twentieth-century, electronic-corpora-based translation studies were mainly initiated by Baker (1993), which began to gain prominence as significant evidential support from a wide range of corpus-based research on translation universals (Baker 1996, Laviosa 1998a). Baker (1996) argues that translation is neither a target language nor a source language in that a translational language has its inherent language properties generated by both source and target languages. She proposes that the universality of translation usually exists in non-native writers' texts caused by unintended by-products of the manipulative mechanism rather than interaction with L1. Several scholars with various ideas have boosted the conceptualization of translation universals (e.g., Baker 1993, 1996, Chesterman 2004b, 2010, Grabowski 2012, Laviosa 2002, Olohan 2004, Xiao 2010)[4], two of which most prominent translationese features pertain to the hypotheses of simplification and explicitation (e.g., Baker 1996, Laviosa 1998a, Lee 2018, 2019, Malmkjar 2012, Mauranen 2007, McEnery and Xiao 2007).

Simplification has become the focus of attention in translationese studies, comprising a good number of linguistic determinants. In the simplification hypothesis, function words are related to the notion of lexical density. The hypothesis of explicitation concerns the matter of readability intertwined with cognitive load. Nevertheless, the cold fact here is that function words as operating variables cause conflicts in these two sub-hypotheses (simplification and explicitation) with significant respect that these variables are not meticulously operated using a unified analytic model. In most prior studies, statistical findings, as yet, have been very much fluctuated research by research so that the widening gulf between findings is derived for every domain, thus being inconsistent with the conventional TU propositions (see Hu, Xiao and Hardie 2019, Kuo 2019, Goh and Lee 2016, Lee 2018, 2019). The simplification hypothesis postulates that the proportion of content words is premised to be small in the non-native writers' translated texts due to their confined level of vocabulary used so that the frequency of function words would appear relatively higher than that of native writer's original texts. Conversely, the explicitation hypothesis assumes that texts are typically written with further details to ease readability and lower readers' cognitive load, thus ending up with sentences longer and more complicated with the excessive use of function words.

As has been said, the proportion of function words is established as an operational variable for translationese measures in both simplification and explicitation hypotheses, which are the most studied but are based on still contradictory premises. Of the TU research conducted to date, most cases have adopted the 'total token' of function words as a TU parameter, whereby very few of the sub-categorized function words have been treated individually. In the translation-related field, most prior studies conducted to date, specifically not until the extensive employment of the machine learning techniques to this field, have adopted the "all-token" approach to function words as a whole (see Chen 2006, Illisei, Inkpen, Pastor and Ruslan 2010, Goh and Lee 2016, Goh, Lee and Kim 2016, Lee 2017, 2018, 2019, Rabinovich and Winter 2015, Xiao 2010). Using Nation's (2013) list of function words, Lee (2017, 2018) conducted a vast array of research into translationese using different monolingual corpora. Primary research queries include testing the hypotheses of translationese using the variables of all-token function words, mean sentence length, lexical density, bottom frequency words, connectives, and lexical bundles.

[4] Some representative constructs include Blum-Kulka's hypothesis (1986), Gellerstam's translationese (1986, 1996), Toury's interlanguage (1979) along with laws of growing standardization and growing interference (1995), Laviosa's core pattern (1998b), and Chesterman's textual fit (2004) along with general patterns and tendencies (2010).

In comparison, regarding the recent studies that have paid attention to function words at the discrete level, Baroni and Bernardini (2006) discussed personal pronouns in the aspect of translationese deploying machine learning methods. Through automatic classification algorithms, Koppel and Ordan (2011) employed 300 individual function words as operating factors to distinguish translated texts from their source languages and reported 92.7% accuracy. Xiao and Dai (2014) compared the composition of both content and function words in the context of Chinese translations, and they found out that auxiliaries, pronouns, prepositions, numerals, and conjunctions are more frequently used in translated texts than non-translated original texts, along with the most significant portion of auxiliaries in translated texts. In the aspects of simplification, explicitation, normalization, and interference in TU hypotheses, Volansky, Ordan and Wintner (2015) also conducted the accuracy tests of text classification using contextual, textual, and functional factors and reported only 77% accuracy level when operated 14 different types of function words as variables. In association with Biber's (1988) multi-dimensional approach, Hu, Xiao, and Hardie (2019) examined 96 linguistic features (not only confined to grammatical functions but also covered textual features) and reported that of the function words category, all preposition tokens, demonstratives, and all article tokens were found to be overrepresented in translational English. Using the variables of function words in Chinese contexts, Kuo (2019) investigated the translationese features of machine-translated texts and found out Chinese conjunctions and adverbials were evident in translated texts.

All told, we have so far looked at prior empirical findings by which function words were adopted as an operating parameter. These prior studies' results were not all consistent about the validity of function words. Some studies argue that having a higher portion of function words is a universal trait that non-native writers commonly share but not what native writers' texts do, while other studies provide strong criticisms against the TU assumptions. The TU hypotheses have conventionally adopted the 'total tokens' approach to function words as analytical measures, but the problem lies in that the results of such methods are still lost in most cases. In that regard, to bridge the gap in the research into L2 writers' texts in Korea, variables of function words as translationese parameters need further exploration for their validation because both lexical density and readability measures are influenced by the portion of function words, thus affecting the feasibility of the simplification and explicitation hypotheses. Hinged on the claim that cognitive processes in writing can be projectable by using the parameter of function words (Pennebaker 2011), the present study will critically examine function words as translationese markers by adopting those variables in both conventional (counting total tokens as a whole) and modified (counting by sub-categories in part) methods. Such research endeavor will give us clues about which parameters of function words make non-native writers' texts separate from native writers' texts and will eventually lead us to tangible arguments. Wherefore, bearing the problem statement in mind, the debate alluded to above leads to the following research questions:

(1) When the variables of function words are operated in a conventional analytical model (i.e., the "all-token approach to function words as a whole), are the translationese manifestations markedly observable in non-native writers' L2 English?
(2) When the subset function words are implemented in a modified analytical model (i.e., the discrete-level approach to function words in part), are any gaps detectable between native writers' L1 English and non-native writers' L2 English?

3. Methods

3.1 Corpus Data

The primary corpus resources were taken from the Comparable Corpora of English Research Abstracts

(CCERA 2.0) to pursue the proposed research aims.[5] The CCERA as monolingual corpora encompasses over 2,000 abstracts from two English-related disciplines: English linguistics and English literature. The CCERA comprises native writers' L1 English and non-native writers' L2 English in both disciplines. Korean scholars from non-Anglophone institutions created the non-native writers' L2 English texts, and these texts are divided into two different types of sub-corpora: one is from those who wrote journal articles and abstracts both in English, and the other is from those who produced articles in Korean and abstracts in English. For the L2 English portion, this study will only use the ones from the former groups in each discipline, and the sub-corpora of L2 English will be labeled as NNW to represent non-native writers' L2 English in both domains. The other sub-corpora of American scholars' English abstracts will be titled NW to mean native writers' original L2 English. By setting two comparable sets (i.e., NW versus NNW in English linguistics and NW versus NNW in English literature), this study will attempt to divulge whether any hallmarks of translationese detected can be considered universal or domain-specific features. The following table briefly shows part of the CCERA investigated for this study.

Table 1. Corpus Design and Textual Statistics

Domain	Sub-Corpus	Text Type	Abstract (#)	Token (#)	Type (#)
English Linguistics	Native Writers (NW)	L1 English	600	105,535	7,594
	Non-Native Writers (NNW)	L2 English	603	106,545	5,898
	Sub Total	•	1,203	212,080	13,492
English Literature	Native Writers (NW)	L1 English	530	106,851	9,743
	Non-Native Writers (NNW)	L2 English	435	105,769	9,086
	Sub Total	•	965	212,620	18,829
	Total	•	2,168	424,700	32,321

3.2 Data Processing and Variable Selection

The ultimate reason for using function words as primary variables (either as a whole or in part) in this study is very deeply related to the three essential rationales. The first one lies in that it is difficult for writers to consciously control the use of function words during the cognitive process of L2 writing. Another rationale is that the use of content words can be affected by choice of registers or topical issues, which would not allow us to have a complete understanding of L2 writing. The last but not most minor rationale is that function words are the only accessible way to the secrecy in cognitive processes in L2 writing. By comparing the composition of function words used in native writers' L1 and non-native writers' L2 writings using the operating parameters of translation universals, this study is fundamentally aimed to revisit translationese as a marker of "non-nativeness" and, more importantly, to discuss the "universality" and "domain-dependency" of translationese in L2 writing.

Several text processing tools were utilized in a complementary manner to tackle the research aims, most of which was the Linguistic Inquiry and Word Count (LIWC) program. The LIWC is a language analysis module that allows researchers to analyze function words and other pertinent grammatical components to interpret text writers' psychological and cognitive processes. Based on approximately 90 output variables, the LIWC module is designed to capture people's language use behaviors from the psychological, linguistic, cognitive, and social dimensions. The module categorizes tokens running across the machine-readable target text files against built-in, on-board dictionaries arranged by multidisciplinary-defined ranges.

[5] Further descriptions and detailed information regarding the self-built CCERA 2.0 can be found in Lee (2017) for CCERA 1.0 and Lee (2018) for CCERA 2.0. The source list where the journal abstracts for the CCERA 2.0 was extracted can be found in Lee (2017).

The LIWC program has been reporting successful performances with the word-count-based interpretation about various psychological and cognitive properties. The updated version of the LIWC2015[6] (Pennebaker, Boyd, Jordan, and Blackburn 2015) text processor comprises approximately 6,400 dictionary words and word stems, encompassing various text types leading to personal writing, natural speech, professional/scientific writing, commercial writing, novels, and many more. Moreover, it usually detects over 86% of the words in individuals' written and speech samples with high reliability. Based on 181,000 text files representing approximately 250 million tokens from numerous corpora produced by over 80,000 writers, the LIWC2015 module computes the portion of total target words (i.e., words for study) that match each of the dictionary categories and provides it in a percentage unit.

During the initial phase of analysis, all the 90 variables set as defaults of the LIWC were carefully screened to judge if those variables play any explicit or implicit roles as function words. Then, only 13 variables out of 90 were chosen to survive. Some of the variables were re-categorized into a new group to meet the needs of the research goals. The 13 variables were classified into two categories (i.e., the all-token function words and quasi-function words). The category of function words was then divided into ten different subsets: function words as a whole (491), the all-token pronouns as a whole (153), personal pronouns (93), impersonal pronouns (59), articles (3), prepositions (74), auxiliary verbs (141), intensifiers (140), conjunctions (43), and negations (62); the category of quasi-function words was also split into three separate subsets such as interrogatives (48), quantifiers (77), and numbers (36). The numbers in the parentheses indicate the types of encoded variables selected for each category. Table 2 provides the statistical information of operational variables adopted for this study.

Table 2. The Statistical Information of Output Variables

Category/Variables	Scale Word Count	Internal Consistency		Professional Writing Means	Grand Means	Means SDs
		Uncorrected **α**	Corrected **α**			
Conventional Variable: Function Words as a Whole						
Function Words_All	492	0.05	0.24	42.39	51.87	5.13
Subset Variables: Function Words in Part						
Pronouns_All	153	0.25	0.67	7.41	15.22	3.61
Personal Pronouns	93	0.2	0.61	3.56	9.95	3.02
Impersonal Pronouns	59	0.28	0.71	3.84	5.26	1.62
Articles	3	0.05	0.23	9.08	6.51	1.79
Prepositions	74	0.04	0.18	14.27	12.93	2.11
Auxiliary Verbs	141	0.16	0.54	5.11	8.53	2.04
Intensifiers	140	0.43	0.82	2.76	5.27	1.61
Conjunctions	43	0.14	0.5	4.85	5.9	1.57
Negations	62	0.29	0.71	0.62	1.66	0.86
Additional Variables: Quasi-Function Words						
Interrogatives	48	0.18	0.57	1.26	1.61	0.76
Quantifiers	77	0.23	0.64	1.94	2.02	0.83
Numbers	36	0.45	0.83	3.55	2.12	2.07

For the interpretation of computed values, in this study, function words as a whole and in part will be treated as a conventional parameter of translationese just as the way conventional TU approaches take. The simplification hypothesis of translation universals posit that more function words will be detected in translated English texts than

[6] For further detailed description of the program, refer to the LIWC manual titled as the Development and Psychometric Properties of LIWC 2015 (Pennebaker, Boyd, Jordan and Blackburn 2015).

in original English utterances. Such premise that translated texts may contain higher portions of grammatical words but lower lexical words typically entail a lower lexical density with less heavy information load (Goh and Lee 2016, Lee 2017, 2018, 2019). Even though the study corpora composed of Korean scholars' L2 writings are not "literally translated," it is premised here that L2 writings are quasi-translated texts, hinged on Cook's (1992: 558) assertion that "L1 and L2 share the same mental lexicon so that L2 processing cannot be cut off from L1."

By taking inductive analyses of language use to discern any discrepancies between native writers' L1 English and non-native writers' L2 English, all the 13 subset dependent variables in each category were thoroughly observed. Then, initial findings from both disciplines were co-compared so that universal, robust makers of translationese could be finally drawn. The subset dependent variables operated in this study are based on the internal reliability and validity provided by the LIWC (Tausczik and Pennebaker 2010). The following matrix in Table 2 outlines the statistical information of the output variables adopted in this study.

A normality test was conducted using the F-Test Two-Sample for Variances analysis tool to evaluate statistical significance between sub-corpora sets. The two-sample F-test performs to compare two population variances and judges each dataset's distribution by determining the shape of distribution as either homogeneity- or heterogeneity-centered. When a *p*-value was less than 0.05, an unequal variance was assumed. Likewise, when a *p*-value exceeded 0.05, the homogeneity of variance was taken. Then, dependent upon the result of a normality test in each variable, a two-sample t-test was selectively performed to evaluate the two sub-corpora differences in each domain. A homoscedastic t-Test was taken for equal variances, and a heteroscedastic t-Test was adopted for unequal variances. Statistical significance was declared when a *p*-value was lower than 0.05. Marginal significance was set at a *p*-value between 0.05 and 0.1. All the statistical analyses were operated using the Analysis ToolPak built in the Microsoft 365 statistics suite and the Compleat Lexical Tutor, an online-based text processing freeware.

4. Results

4.1 A Conventional Model: Function Words as a Whole

This study has taken the first attempt to revisit function words' feasibility as a marker of translationese in non-native writers' L2 English texts by employing a two-fold analysis. The first tier was to take the "all-token" analytic method as an operating variable as conventionally implemented in the descriptive translation studies. The second tier was to examine individual sub-categories of subset function words and additional quasi-function words using a modified analytic model on a discrete level. According to the conventional postulation of translation universals, it is assumed that function words are likely to be more salient in translated texts than in original texts, and this study will adopt the same framework to report the analysis results by stating which sub-corpora have a higher value of variables than their counterpart.

When the all-token method of function words was taken as a whole, the mean differences between the two corpus pairs in both disciplines indicated significant differences but not in the same way. In the English linguistics discipline, native writers' L1 English (NW) corpus was shown to have more function words than non-native writers' L2 English (NNW) corpus, which in turn was found to be statistically significant ($t(1,201) = 3.473$, $p < .001$) but opposed to the conventional proposition of the TU hypothesis.

On the other hand, a distinctive pattern was observed in the English literature domain. That is, non-native writers' L2 (NNW) corpus had a higher value of function words than that of the native writers' L1 (NW) corpus with statistical significance ($t(963) = -3.715$, $p < .001$), conforming to the syntactic simplification in the hypotheses of translation universals. The summary output of the analysis results is outlined in Table 3 and Table 4.

Table 3. Group Statistics: English Linguistics

English Linguistics Variables	Normality Tests			Test Values					
	F	*p*		Group	*N*	*M*	*SD*	*t*	*p*
Conventional Variable:	***Function Words as a Whole***								
Function Words_All	0.782	0.001	$p < 0.05$	**NW**	600	**42.753**	17.757	3.473	**0.000*****
				NNW	603	41.852	22.711		
Subset Variables:	***Function Words in Part***								
Pronouns_All	1.038	0.325	$p > 0.05$	**NW**	600	**5.833**	5.391	6.286	**0.000*****
				NNW	603	4.999	5.196		
Personal Pronouns	0.881	0.061	$p > 0.05$	NW	600	1.316	1.601	0.720	0.236
				NNW	603	1.262	1.818		
Impersonal Pronouns	1.203	0.012	$p < 0.05$	**NW**	600	**4.516**	3.615	7.432	**0.000*****
				NNW	603	3.737	3.005		
Articles	0.827	0.010	$p < 0.05$	NW	600	9.166	7.019	0.639	0.261
				NNW	603	9.064	8.487		
Prepositions	1.006	0.470	$p > 0.05$	NW	600	16.070	6.554	1.249	0.106
				NNW	603	15.886	6.514		
Auxiliary Verbs	1.167	0.029	$p < 0.05$	NW	600	4.604	4.459	-2.105	**0.018***
				NNW	603	**4.851**	3.821		
Intensifiers	1.297	0.001	$p < 0.05$	NW	600	2.184	1.811	-2.222	**0.013***
				NNW	603	**2.368**	2.349		
Conjunctions	0.911	0.128	$p > 0.05$	NW	600	5.891	5.971	-0.685	0.247
				NNW	603	3.877	4.254		
Negations	1.038	0.323	$p > 0.05$	NW	600	0.671	1.335	0.205	0.419
				NNW	603	0.658	1.286		

*Statistical significance is set at ***$p < .001$, **$p < .01$, and *$p < .05$.

4.2 A Modified Model: Function Words in Part

As for the second tier of the analyses, nine subsets of function words (i.e., total pronouns, personal pronouns, impersonal pronouns, articles, prepositions, auxiliary verbs, intensifiers, conjunctions, and negations) were additionally operated as discrete-level variables from grammatical categories, although such approach has been operated extensively yet, neither in translation studies nor in the research into L2 writing. By employing function words at a discrete level by their sub-categories, the aspects of "universality" in translationese were intended to be reconsidered.

When individual variables of function words were examined as translationese indices in the English linguistics domain, four out of nine variables (total pronouns, impersonal pronouns, auxiliary verbs, and intensifiers) were found to show statistically significant differences. Even with their statistical significance, only half of the variables (auxiliary verbs and intensifiers) were further pervasive in the non-native writers' NNW corpus than in the native writers' NW corpus. Non-native writers' NNW texts had a more significant percentage of auxiliary verbs than that of native writers' NW texts with statistical significance ($t(1,201) = -2.105$, $p < .05$), meaning that such result is consistent with the TU presupposition. Another variable that indicated statistical significance in non-native writers' dominance was observed in the use of intensifiers. Similarly, the values of intensifiers were higher in the NNW than in the NW corpora with statistical significance ($t(1,201) = -2.222$, $p < .05$), which in turn falls in with the TU hypothesis.

Table 4. Group Statistics: English Literature

English Literature Variables	Normality Tests			Test Values					
	F	*p*		Group	*N*	*M*	*SD*	*t*	*p*
Conventional Variable:	***Function Words as a Whole***								
Function Words_All	0.838	0.027	$p < 0.05$	NW	530	42.911	16.091	-3.715	**0.000*****
				NNW	435	**43.925**	19.192		
Subset Variables:	***Function Words in Part***								
Pronouns_All	0.957	0.313	$p > 0.05$	NW	530	6.018	5.580	-2.974	**0.002****
				NNW	435	**6.477**	5.833		
Personal Pronouns	0.790	0.005	$p < 0.05$	NW	530	2.131	2.959	-4.466	**0.000*****
				NNW	435	**2.663**	3.743		
Impersonal Pronouns	0.998	0.488	$p > 0.05$	NW	530	3.886	2.521	0.704	0.241
				NNW	435	3.813	2.527		
Articles	0.899	0.121	$p > 0.05$	NW	530	9.782	6.296	-1.603	0.055
				NNW	435	10.049	7.006		
Prepositions	0.995	0.479	$p > 0.05$	**NW**	530	**16.338**	4.961	1.810	**0.035***
				NNW	435	16.077	4.984		
Auxiliary Verbs	1.161	0.052	$p > 0.05$	NW	530	3.477	3.169	-5.017	**0.000*****
				NNW	435	**4.036**	2.729		
Intensifiers	1.121	0.108	$p > 0.05$	NW	530	2.309	1.393	1.261	0.104
				NNW	435	2.215	1.242		
Conjunctions	1.052	0.291	$p > 0.05$	NW	530	6.490	3.557	-1.501	0.067
				NNW	435	**6.672**	3.381		
Negations	0.976	0.394	$p > 0.05$	NW	530	0.450	0.343	-4.547	**0.000*****
				NNW	435	**0.623**	0.352		

*Statistical significance is set at ***$p < .001$, **$p < .01$, and * $p < .05$.

Meanwhile, native writers' prevalent features were noticeable in the last half (total pronouns and impersonal pronouns) of the four variables. The feature of total pronouns was found to be more apparent in native writers' NW corpus rather than the non-native writers' NNW corpus, being inconsistent with the conventional TU hypothesis ($t(1,201) = 6.286$, $p < .001$). Likewise, the variable of impersonal pronouns was also further marked in the NNW corpus than in the NW corpus, resulting in statistical significance ($t(1,201) = 7.432$, $p < .001$).

In the domain of English literature, five out of nine subset variables (total pronouns, personal pronouns, prepositions, auxiliary verbs, and negations) were turned out to be statistically significant. Two of them showed only a marginal significance as shown in the variables of articles ($t(963) = -1.603$, $p = 0.055$) and conjunctions ($t(963) = -1.501$, $p = .067$), in which the NNW corpus was more prominent behavior in the use of both variables as compared to the NW corpus. On the other hand, of the five variables as mentioned above with statistical significance, non-native writers' NNW corpus was more dominant than the native writers' NW corpus in the use of total tokens of pronouns ($t(963) = -2.974$, $p < .01$), impersonal pronouns ($t(963) = -4.466$, $p < .001$), auxiliary verbs ($t(963) = -5.017$, $p < .001$), and finally negations ($t(963) = -4.547$, $p < .001$), which in turn demonstrated consistency with the TU presupposition.

The only similar finding observable as in the other domain was the salient feature of auxiliary verbs. As compared to the NW corpus, a higher portion of auxiliary verbs was noticeable in the non-native writers' NNW corpus, indicating the only "universal" variable to fit in the simplification hypothesis of translation universals in both scholarly disciplines. Even with its statistical significance, however, the last variable (prepositions) was found to be more salient in the NW texts rather than the NNW corpus, demonstrating a contrary result to the TU

assumption ($t(963) = 1.810$, $p < .05$). Table 3 delineates the summary output of both conventional and discrete-level subset variables in the English linguistics domain and Table 4 for English literature, respectively.

4.3 Additional Variables: Quasi-Function Words

For an additional tier of analysis, the category of quasi-function words was investigated as well. The LIWC2015 module has a default built-in dictionary setting that categorizes six variables of English grammatical items (common verbs, common adjectives, comparisons, interrogatives, quantifiers, and numbers) into the subsets titled "other grammar." However, in this study, only three grammatical items, which seem to take the least content-based roles, were picked out to be placed in a newly coined category called quasi-function words.

Of the three variables selected for the last-tier analysis, it was found that all the three variables (interrogatives, numbers, and quantifiers) in the English linguistics domain and two factors (quantifiers and numbers) from the English literature group demonstrated overriding attributes in those variables, respectively, showing statistically significant results but in a distinctive pattern in each discipline. To be specific, in the English linguistics discipline, native writers' L1 English NW corpus took up for a higher portion of interrogatives than that of the non-native writers' L2 English NNW corpus, resulting in statistically significant differences in mean values ($t(1,201) = 3.241$, $p < .001$). Contrariwise, the non-native writers' NNW corpus outperformed in the use of two variables distinctively from the native writers' NN corpus, which again was found to be statistically significant in the use of quantifiers ($t(1,201) = -3.195$, $p < .001$) as well as the use of numbers ($t(1,201) = -3.332$, $p < .001$), still complying with the TU suppositions. The specific results are mapped at the bottom of Table 5.

Table 5. Group Statistics: English Linguistics

English Linguistics Variables	Normality Tests			Test Values					
	F	*p*		Group	*N*	*M*	*SD*	*t*	*p*
Additional Variables:	***Quasi-Function Words***								
Interrogatives	1.179	0.022	$p < 0.05$	**NW**	600	**0.977**	0.894	3.241	**0.001*****
				NNW	603	0.807	0.758		
Quantifiers	0.716	0.000	$p > 0.05$	NW	600	1.579	1.342	-3.195	**0.001*****
				NNW	603	**1.812**	1.875		
Numbers	0.861	0.034	$p < 0.05$	NW	600	1.904	2.621	-3.332	**0.000*****
				NNW	603	**2.247**	3.740		

*Statistical significance is set at ***$p < .001$, **$p < .01$, and *$p < .05$.

In the English literature discipline, the same trend was not discernible, however. The native writers' L1 English NW corpus overwhelmed the non-native writers' L2 English NNW corpus in the values of quantifiers ($t(963) = 6.084$, $p < .001$) and the parameter of numbers ($t(963) = 7{,}662$, $p < .001$) with statistical significance. Contrary to the other domain, the NNW corpus did not represent these variables' pervasive use, so both results did not support translation universals' simplification. Table 6 shows the overall results concerning the quasi-function words tested in English literature.

Table 6. Group Statistics: English Literature

English Literature Variables	Normality Tests			Test Values					
	F	*p*		Group	*N*	*M*	*SD*	*t*	*p*
Additional Variables:	***Quasi-Function Words***								
Interrogatives	0.967	0.355	$p > 0.05$	NW	530	1.155	0.641	-0.207	0.418
				NNW	435	1.166	0.663		
Quantifiers	1.620	0.000	$p < 0.05$	**NW**	530	**1.473**	1.044	6.084	**0.000*****
				NNW	435	1.116	0.645		
Numbers	1.732	0.000	$p < 0.05$	**NW**	530	**1.763**	2.395	7.662	**0.000*****
				NNW	435	1.091	1.383		

*Statistical significance is set at ***$p < .001$, **$p < .01$, and *$p < .05$.

5. Discussion

For the evaluation of function words' feasibility as markers of translationese, this study mixed two different analytical models by implementing function words in one whole unit and the nine subset categories of function words in part along with three additional categories of grammatical words at the discrete level. Based on the outputs obtained from the two-fold analyses, the forthcoming sections will therefore discuss the extent to which research aims were pursuable, drawn by the results and subsequent findings to understand the identity of non-nativeness.

5.1 As a Whole or In Part

Regarding the use of the total tokens of function words in a conventional manner, the two disciplines have emerged with completely contrary results to each other. A more significant percentage of function words in the English linguistics discipline was detectable as the native writers' L1 English attribute, whereas non-native writers' L2 English was more salient using function words in the English literature domain. These results seem to be not surprising but predictable, backed by prior studies (Goh and Lee 2016, Lee 2017, 2018, 2019). One thing notable here is that such results collide with the matter of "universality" in the conventional TU hypotheses, as it was turned out to be valid only in one specific domain (English literature) where native writers were overwhelmed by non-native writers regarding the variable of the all-token function words as a whole.

One plausible idea to understand such cases would be that most prior studies in translation universals were mostly literary-centric. Even if research abstracts belong to academic prose, it can be easily deduced that those texts from the English literature discipline would probably be leaned towards literary genres and registers distinctive to literature. More notably, as affirmed in Biber's (1995, 1998) argument, words in the non-native writers' L2 English corpus may have been clustered by their grammatical functions, thereby affecting the frequency of function words caused by register variations between the two domains. The second cause is related to the way the list of function words is created. In most cases, the "all-token" lists of function words utilized as a stop list in text processing phases tend to be varied research by research up to which sources those lists are extracted. Some studies often include numbers, quantifiers, grammatical expletives, degree adverbs, and intensifiers in their lists as necessary, while others may not. Still, when the studies' results are reported, the specific categories of lists are not likely to be available. Finally, the class of connectives is one of the representative variables used to test the hypothesis of explicitation, which collides with the proposition of simplification.

As can be seen from prior studies, it was not very safe to validate the TU hypotheses by relying solely on the percentage of the total tokens of function words as a whole. As noted several times earlier in previous findings, the only way to challenge the operational definition of the conventional TU hypotheses was to take an alternative

by turning our eyes to the high-frequency function words (e.g., function words ranked up to top 5, 10, 20, …) as a "modified" parameter (see Goh and Lee 2016, Lee 2017, 2018, 2019). The problem that arises here is that high-frequency function words typically end up being positioned at the top ranks like other high-frequency words, which ultimately will limit their assigned role as a determinant of lexical density, rendering the debate continued over the conflicting properties as a TU predictor.

Thus, if function words are to be adopted as a TU predictor as in a conventional way, the all-token approach to the variable of function words as a whole unit should no longer be considered a potential candidate parameter for validating the TU hypotheses. Instead, it is necessary to re-establish the definition of the "universality" of translationese manifestations as long as we consider lexical density as part of the TU measure calculated based on the total token of function words.

Meanwhile, it was noteworthy that analyzing the unit of individual subset categories of function words yields more robust results than operated with the total tokens of function words in a whole unit. As for the second-tier of the analyses, nine subset variables of function words were additionally operated at a discrete level from grammatical categories. Such an approach has been operated extensively yet, neither in translation studies nor in the research into L2 writing. As illustrated in the mark "●" in Table 7, it was shown that the TU hypothesis is valid in both groups. Only the factor of auxiliary verbs had proven to be justifiable in both academic disciplines when subset variables were tested, thus seen as a reconfirmation of the existing TU proposition that non-native writers' English is not homogeneous with that of native writers.

By taking the modified method rather than the all-token method, it was deduced that auxiliary verbs have the potential as a "robust" translationese indicator. Several empirical findings have supported similar results to what has been found here regarding the excessive use of auxiliary verbs (see Kuo 2019, Xiao and Dai 2014).

Also, projected in the mark "◑" in Table 7, it was demonstrated that though not universal, the potentiality of domain-specific translationese had better be reweighed so that future research can be directed targeting those half-marked variables to be tested.

Though predictable from prior studies, what has been derived here can still be surprising and intriguing in many different aspects since the analysis results may give us a clue why the all-token function words have been such an unstable TU parameter. Table 7 shows the output of the modified model only with the significant variables.

Table 7. Variable Comparison: As a Whole vs. In Part

Variable Type	English Linguistics	TU Verified	English Literature	TU Verified
Conventional Variable				
Function Words_All	x	x	**NW < NNW**	◑
Subset Variables				
Pronouns_All	x	x	**NW < NNW**	◑
Personal Pronouns	x	x	**NW < NNW**	◑
Impersonal Pronouns	x	x	x	x
Prepositions	x	x	x	x
Auxiliary Verbs	**NW < NNW**	●	**NW < NNW**	●
Intensifiers	**NW < NNW**	◑	x	x
Negations	x	x	**NW < NNW**	◑

* The pairs marked with the inequality sign "<" indicate that the NNW group has more values of function words than those of the NW group, complying with the TU simplification hypothesis. The mark "●" indicates that the conventional TU hypothesis is valid in both domains. The mark "◑" represents that the given slot variable satisfied only a single domain, either English linguistics or English literature.

5.2 Domain-Specific Translationese

Contrary to the results drawn using the all-token approach to function words, a majority of the subset variables were turned out to be statistically significant irrespective of the conventional TU presupposition, a result that either non-native writers' L2 texts outperformed their counterparts or vice versa, except for the only universal variable of auxiliary verbs discussed in Section 5.1. In specific, several factors were marked as consistent with the TU assumption but not in all the selected variables.

One of the most intriguing findings here is that by applying a discrete-level modified approach to function words, the intrinsic differences present in native and non-native writers' texts were perceived by the disciplinary distance between English linguistics and English literature corpora. The non-native writers in the English literature discipline overwhelmed their counterparts by the seven subset variables (i.e., five with statistical significance and two with marginal significance) in the function words category.

In comparison, non-native L2 English subdued native writers' L1 English only in two subset variables. Such results are thus seen as a reconfirmation of the existing TU proposition's unstable conceptualization. Instead, it was evidenced that the English linguistics domain's non-native writers' L2 English is not homogeneous with that of the English literature domain. It can be construed that the subset variables are worth considering as markers of prospective "domain-specific" translationese. The conventional TU postulation that non-native writers use more tokens of function words than native writers do across disciplines, thus, should be rejected. Instead, adopting a modified analytic model to function words (i.e., utilizing subset variables) as domain-specific translationese markers-may be further accessible.

Table 8. Feature Comparison: Domain-Specific Translationese

Variable	English Linguistics		English Literature	
	Native Writers	Non-Native Writers	Native Writers	Non-Native Writers
Universal Feature				
Auxiliary Verbs	x	**NNW**	x	**NNW**
Domain-Dependent Features				
Function Words_All	NW	x	x	**NNW**
Pronouns_All	NW	x	x	**NNW**
Personal Pronouns	x	x	x	**NNW**
Impersonal Pronouns	NW	x	x	x
Prepositions	x	x	NW	x
Intensifiers	x	**NNW**	x	x
Negations	x	x	x	**NNW**

* The marks of **NW** and **NNW** in each discipline column's slots indicate the dominant group (i.e., native writers' L1 English or non-native writers' L2 English), which holds a higher portion of specific function words than its counterpart.

Like the output produced by the subset variables of function words, three additional grammatical categories, titled quasi-function words, showed statistically significant differences among paired corpora in an ambivalent way between domains, specifically with two variables. In the English linguistics discipline, the additional variables of numbers and quantifiers were outperformed in non-native writers' L2 English, complying with the TU assumption, where those variables were underperformed in the other domain. The interrogative variable was only influential in native writers' L1 English in the English linguistics corpus, but no significant findings were not detectable in their counterpart discipline. When interpreting these results, what has to be taken into account is that domain-specific topics and registers might have been provoking to a certain extent with overpopulated function words in the English linguistics discipline. Even though quantifiers are often classified into a determiner category, they are distinctive to numerals indicating precise values or amounts. Both numbers and quantifiers

describe quantity, and the English linguistics domain would likely outrun in the use of these quantity words compared to the English literature domain. As noted earlier, these quasi-function words can also be considered domain-specific predictors of translationese as opposite results have been drawn up between both domains. Table 9 visualizes the output of the quasi-function words category.

Table 9. Quasi-Function Words as Translationese Markers

Variable	English Linguistics		English Literature	
Quasi-Function Words				
Interrogatives	NW	x	x	x
Quantifiers	x	**NNW**	NW	x
Numbers	x	**NNW**	NW	x

***** The marks of **NW** and **NNW** given in each discipline column's slots indicate the dominant group (i.e., native writers' L1 English or non-native writers' L2 English) that holds a higher portion of quasi-function words than its counterpart.

As raised by neighboring disciplines (Pennebaker 2015, Pennebaker and Graybeal 2001), non-native writers can purposely control their own content words for writers' voices or preferred lexical words. That means that their consciousness can be significantly influenced when choosing particular content words to use when generating content-specific texts, for example. As non-native professional writers and scholars, they must be very familiar with big words or registers pertinent to their research fields and would have a command of manipulating such content words as to their voices or preferences, as necessary. Non-native writers with high proficiency can also consciously increase content words' weight during writing's cognitive processes. As many prominent researchers (Biber 1988, Laufer and Nation 1995, 2014, Pennebaker 2003, 2007) claimed, function words' value is inversely proportional to specific nouns, including subordinate clauses. The higher the frequency of function words, the lower that of specific nouns, meaning that function words influenced how content words are used. Though each domain drove conflicting results, it can still be speculated that non-native writers from both disciplines may lack "domain-specific" requirements. The notion of domain-specific requirements regards domain-dependent academic writing expertise in the L2, rather than only being confined to general writing competence. In L2 writing, domain-specific expertise is often discussed using the term "epistemic writing." As depicted in cognitive models of writing (Hayes and Flower, 1980), writing involves the modes of elaborating and knowledge-constituting, not only limited to generating functions (Bangert-Drowns, Hurley and Wilkinson, 2004, Bereiter and Scardamalia 1987). In simple terms, epistemic writing is often understood as "writing-to-learn" (Keys 1999). As such, expert L2 writers should be equipped with domain-specific knowledge-structuring at the semantic level, not only constricted to syntactic and lexical levels of writing competencies.

Overall, although not substantiated the "universality" of translationese across domains, it can be speculated that function words still DO matter. That is because function words provide us with clues to questions about the gulf between nativeness and non-nativeness. Bridging the gap in using function words between native L1 and non-native L2 writers will be one of the overarching matters in identifying accessible predictors to spot translationese regardless of either "universal" or "domain-specific."

6. Conclusion

The present research sought to provide meaningful additions to those limited in explaining function words as markers of translationese by intermingling conventional and modified explanations. The parameter of "all-token" function words has been one of the most unstable TU indicators in conventional translation studies. Furthermore, most prior studies using the all-token approach to function words have been confined to giving quantitative

explanations only by using their numerical values and did not fully explain how specific function words varied in translationese predictors' roles. Therefore, this study adopted both a conventional analytic model (i.e., all-token function words) and a modified analytic model (i.e., subset function words) as principal variables along with three additional variables (i.e., quasi-function words) that could never be clarified to date.

According to the hypotheses of translation universals, it has been postulated that native writers' texts have a higher rate of lexical density by holding a lower portion of function words but a higher portion of lexical words. The results of this study were contrary to common assumptions of translation universals. All-token function words were more discernible only in English literature but not in the English linguistics discipline. Instead, auxiliary verbs demonstrated a higher predictive and "universal" power as a newly tested translationese marker. Thus, the feature of "universality" in the TU hypotheses should be reconsidered in the sense that eleven out of thirteen variables from the function words and quasi-function word categories were overshadowed by more salient attributes of each domain. In this respect, function words as a marker of "translationese" (i.e., either universal or genre-specific) remain a valid parameter for identifying non-nativeness. Even if it may have some drawbacks, it should not be rejected entirely but reconceptualize translationese' universality to steer future research directions further.

References

Baker, M. 1993. Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis and E. Tognini-Bonelli, eds., *Text and Technology: In Honour of John Sinclair*, 233-250. Amsterdam: John Benjamins.

Baker, M. 1995. Corpora in translation studies: An overview and some suggestions for future research. *Target* 7(2), 223-243.

Baker, M. 1996. Corpus-based translation studies: The challenges that lie ahead. In H. Somers, ed., *Terminology, LSP and Translation*, 175-186. Amsterdam and Philadelphia: John Benjamins.

Baker, M. 2004. A corpus-based view of similarity and difference in translation. *International Journal of Corpus Linguistics* 9(2), 167-193.

Baker, M. 2007. Patterns of idiomaticity in translated vs. non-translated text. *Belgian Journal of Linguistics* 21(1), 11-21.

Baroni, M. and S. Bernardini. 2006. A new approach to the study of translationese: Machine-learning the difference between original and translated text. *Literary and Linguistic Computing* 21(3), 259-274.

Bangert-Drowns, R., M. M. Hurley and B. Wilkinson. 2004. The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research* 74, 29-58.

Bereiter, C. and M. Scardamalia. 1987. *The Psychology of Written Composition*. Hillsdale, NJ: Lawrence Erlbaum.

Bernardini, S. and F. Zanettin. 2004. When is a universal not a universal? Some limits of current corpus-based methodologies for the investigation of translation universals. In A. Mauranen and P. Kujamäki, eds., *Translation Universals: Do They Exist?*, 51-62, Amsterdam: John Benjamins.

Bialystok, E., F. Craik and G. Luk. 2008. Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition* 34(4), 859-873.

Biber, D. 1988. *Variation Across Speech and Writing.* Cambridge: Cambridge University Press.

Biber, D. 1995. *Dimensions of Register Variation: A Cross-linguistic Comparison*. Cambridge: Cambridge University Press.

Blum-Kulka, S. 1986. Shifts of cohesion and coherence in translation. In J. House and S. Blum-Kulka, eds., *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second*

Language Acquisition Studies, 17-35. Tübingen: Gunter Narr.
Campbell, R. S. and J. W. Pennebaker. 2003. The secret life of pronouns: Flexibility in writing style and physical health. *Psychological Science* 14, 60-65.
Cao, Y. and R. Xiao. 2013. A multi-dimensional contrastive study of English abstracts by native and non-native writers. *Corpora* 8(2), 209-234.
Connor, U. 1990. Linguistic/rhetorical measures for international persuasive student writing. *Research in the Teaching of English* 24(1), 67-87.
Cook, V. 1992. Evidence for multicompetence. *Language Learning* 42(4), 557-591.
Cumming, A. 1990. Metalinguistic and ideational thinking in second language composing. *Written Communication* 7(4), 482-511.
Chen, J. W. 2006. *Explicitation through the Use of Connectives in Translated Chinese: A Corpus-study*. Doctoral dissertation, The University of Manchester, Manchester, England.
Chesterman, A. 2004a. Beyond the particular. In A. Mauranen and P. Kujamaki, eds., *Translation Universals: Do They Exist?*, 33-49. Amsterdam: Benjamins.
Chesterman, A. 2004b. Hypotheses about translation universals. In G. Hansen, K. Malmkjar and D. Gile, eds., *Claims, Changes, and Challenges in Translation Studies*, 1-13. Amsterdam: John Benjamins.
Chesterman, A. 2010. Why study translation universals? In R. Hartama-Heinonen and P. Kukkonen, eds., *Kiasm*, 38-48. Helsinki: University of Helsinki.
Chung, C. K. and J. W. Pennebaker. 2012. Linguistic inquiry and word count (LIWC): Pronounced “Luke,”... and other useful facts. In P. M. McCarthy and C. B. Denecke, eds., *Applied Natural Language Processing: Identification, Investigation, and Resolution*, 206-229. Hershey, PA: IGI Global.
Chambers, J. K., P. Trudgill and N. Schilling¬Estes. 2004. *The Handbook of Language Variation and Change*. London: Blackwell.
Chung, C. K. and J. W. Pennebaker. 2013. Using computerized text analysis to track social processes. In T. Holtgraves, ed., *Handbook of Language and Social Psychology*, 219-23. New York, NY: Oxford.
Chung, C. K. and J. W. Pennebaker. 2007. The psychological functions of function words. In K. Fiedler, eds., *Social Communication*, 343-359. New York: Psychology Press.
Cobb, T. Vocabulary statistics v.3 [computer program]. Accessed 5 January 2021 at https://www.lextutor.ca/stats/
Frawley, W. 1984. Prolegomenon to a theory of translation. In W. Frawley, ed., *Translation: Literary, Linguistic and Philosophical Perspectives*, 159-175. Newark: University of Delaware Press.
Gaspari, F. and S. Bernardini. 2008. *Comparing non-native and translated language: Monolingual Comparable Corpora with a Twist*. Paper presented at the International Symposium on Using Corpora in Contrastive Translation Studies.
Gellerstam, M. 1986. Translationese in Swedish novels translated from English. In L. Wollin and H. Lindquist, eds., *Translation Studies in Scandinavia*, 88-95. Lund: CWK Gleerup.
Gellerstam, M. 1996. Translations as a source for cross-linguistic studies. In K. Aijmer, B. Altenberg and M. Johansson, eds., *Languages in Contrast*, 53-6. Lund: Lund University Press.
Goh, G. Y. 2007. How authentic is the language of Korean middle school English textbooks? A corpus-based analysis. *Korean Journal of English Language and Linguistics* 7(2), 191-210.
Goh, G. Y. and Y. C. Lee. 2016. A corpus-based study of translation universals in English translations of Korean newspaper texts. *Cross-Cultural Studies* 45, 109-143.
Goh, G. Y., Y. C. Lee and D. Y. Kim. 2016. A corpus-based study of translation universals in thesis/dissertation abstracts. *Korean Journal of English Language and Linguistics* 16(4), 819-849.
Grabowski, L. 2012. On translation universals in selected contemporary Polish literary translations. *Studies in Polish Linguistics* 7, 165-183.
Hartley, J. 2002. Notetaking in non-academic settings: A review. *Applied Cognitive Psychology* 16, 559-574.

Hartley, J., E. Sotto and J. W. Pennebaker. 2002. Style and substance in psychology: Are influential articles more readable than less influential ones. *Social Studies of Science* 32, 321¬334.

Hartmann, R. R. K. and F. C. Stork. 1972. *Dictionary of Language and Linguistics*. Essex: Applied Science Publishers.

Harwood, N. 2005. We do not seem to have a theory... the theory I present here attempts to fill this gap: Inclusive and exclusive pronouns in academic writing. *Applied Linguistics* 26(3), 343-75.

Hayes, J. R. and L. Flower. 1980. Identifying the organization of writing processes. In L. W. Gregg and E. R. Steinberg, eds., *Cognitive Processes in Writing: An Interdisciplinary Approach,* 3-30. Hillsdale, NJ: Lawrence Erlbaum.

Heberlein, A. S., R. Adolphs, J. W. Pennebaker and D. Tranel. 2003. Effects of damage to right-hemisphere brain structures on spontaneous emotional and social judgments. *Political Psychology* 24, 705-726.

Hönig, H. G. 1991. Holmes's 'mapping theory' and the landscape of mental translation processes. In M. Kitty, V. Leuven-Zwart and T. Naaijkens, eds., *Translation Studies: State of the Art*, 77-91. Amsterdam-Atlanta: Rodopi.

Hu, X. and R. Xiao. 2014. How different is English translation from native writings of English? A multi-feature statistical model for linguistic variation analysis. In *Papers from the 35th ICAME Conference on Corpus Linguistics, Context and Culture*, 106-107.

Hu, X., R. Xiao and A. Hardie. 2019. How do English translations differ from native English writings? A multi-feature statistical model for linguistic variation analysis. *Corpus Linguistics and Linguistic Theory* 15(2), 347-382.

Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Illisei, I., D. Inkpen, G. C. Pastor and M. Ruslan. 2010. Identification of translationese: A machine learning approach. In *Papers from the 11th International Conference on Computational Linguistics and Intelligent Text Processing,* 503-511.

Kennedy, G. 1988. *An Introduction to Corpus Linguistics*. London: Longman.

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

Kern, R. 1994. The role of mental translation in second language reading. *Studies in Second Language Acquisition* 16, 441-461.

Keys, C. W. 1999. Language as an indicator of meaning generation: An analysis of middle school students' written discourse about scientific investigations. *Journal of Research in Science Teaching* 36, 1044-1061.

Koppel, M. and N. Ordan. 2011. *Translationese and its dialects*. Paper presented at the Association for Computational Linguistics.

Kunilovskaya, M. and E. Lapshinova-Koltunski. 2019. *Translationese features as indicators of quality in English-Russian human translation*. Paper presented at the Association for Computational Linguistics.

Kuo, C. 2019. Function words in statistical machine-translated Chinese and original Chinese: A study into the translationese of machine translation systems. *Digit. Scholarsh. Humanit* 34, 752-771.

Laufer, B. and I. S. P. Nation. 1995. Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics* 16(3), 307-322.

Laviosa, S. 1998a. The corpus-based approach: A new paradigm in translation studies. *Meta: Translators' Journal* 43(4), 474-479.

Laviosa, S. 1998b. Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta: Translators' Journal* 43(4), 557-570.

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

Lee, Y. C. 2017. *The Hallmarks of L2 Writers' Texts Viewed through the Prism of Translation Universals: A Corpus-based Approach to English Research Abstracts of Scholarly Journal Articles*. Doctoral Dissertation, Yonsei University, Seoul, Korea.

Lee, Y. C. 2018. The hallmarks of L2 writing viewed through the prism of translation universals. *Linguistic Research* 35, 171-205.

Lee, Y. C. 2019. Spotting non-nativeness in L2 texts: A statistical approach to translationese. *Studies in English Language and Literature* 45(1), 367-388.

Malmkjær, K. 2005. Norms and nature in translation studies. *Synaps* 16, 13-19.

Malmkjær, K. 2012. *Handbook of Translation Studies*. Amsterdam: John Benjamins.

Mauranen, A. 2003. The corpus of English as lingua franca in academic settings. *TESOL Quarterly* 37(3), 513-527.

Mauranen, A. 2007. Universal tendencies in translation. In G. Anderman and M. Rogers, eds., *Incorporating Corpora: The Linguist and the Translator*, 32-48. Bristol: Multilingual Matters.

McEnery, T. and A. Wilson. 2001. *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.

McEnery, T. and R. Xiao. 2007. Quantifying constructions in English and Chinese: A corpus-based contrastive study. In *Papers from the Corpus Linguistics Conference*, 27-30.

McEnery, T., R. Xiao and Y. Tono. 2006. *Corpus-based Language Studies: An Advanced Resource Book*. London and New York: Routledge.

Newmark, P. 1991. The curse of dogma in translation studies. *Lebende Sprachen* 36(3), 105-108.

Nation, I. S. P. 2013. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

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

Pennebaker, J. and A. Graybeal. 2001. Patterns of natural language use: Disclosure, personality, and social integration. *Current Directions in Psychological Science* 10, 90-93.

Pennebaker, J. W. 2003. The social, linguistic, and health consequences of emotional disclosure. In J. Suls and K. A. Wallston, eds., *Social Psychological Foundations of Health and Illness*, 288-313. Malden: Blackwell Publishing.

Pennebaker, J. W., R. J. Booth and M. E. Francis. 2007. *Linguistic Inquiry and Word Count: A Text Analysis Program*. Austin, TX: LIWC.net.

Pennebaker, J. W. 2011. *The Secret Life of Pronouns: What Our Words Say about Us*. New York: Bloomsbury Press.

Pennebaker, J. W. and C. K. Chung. 2013. Counting little words in big data: The psychology of individuals, communities, culture, and history. In J. Laszlo, J. P. Forgas and O. Vincze, eds., *Social Cognition and Communication,* 173-184. New York: Psychology Press.

Pennebaker, J. W., R. J. Booth, R. L. Boyd and M. E. Francis. 2015. *Linguistic Inquiry and Word Count: LIWC 2015*. Austin, TX: Pennebaker Conglomerates, Inc.

Pennebaker, J. W., R. L. Boyd, K. Jordan and K. G. Blackburn. 2015. *The Development and Psychometric Properties of LIWC 2015.* Austin, TX: Pennebaker Conglomerates, Inc.

Pennebaker, J. W., M. R. Mehl and K. G. Niederhoffer. 2003. Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology* 54, 547-577.

Pennebaker, J. and L. King. 1999. Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology* 77(6), 1296-1312.

Perani, D., D. Stanislas, F. Grassi and J. Mehler. 1996. Brain processing of native and foreign languages. *NeuroReport* 7(15), 2439-2444.

Pym, A. 2008. On Toury's laws of how translators translate. In A. Pym, M. Shlesinger and D. Simeoni, eds., *Beyond Descriptive Translation Studies. Investigations in Homage to Gideon Toury*, 311-328. Amsterdam: John Benjamins.

Qi, D. S. 1998. An inquiry into language-switching in second language composing processes. *The Canadian Modern Language Review* 54(3), 413-435.

Quirk, R., S. Greenbaum, G. Leech and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*.

London: Longman.

Rabinovich, E. and W. Winter. 2015. Unsupervised identification of translationese. *Transactions of the Association for Computational Linguistics* 3, 419-432.

Ransdell, S. and M.-L. Barbier. 2002. An introduction to new directions for research in L2 writing. In S. Ransdell and M.-L. Barbier, eds., *New Directions for Research in L2 Writing*, 1-10. Dordrecht: Kluwer Academic Publishers.

Secara A. 2005. Translation evaluation: A state of the art survey. In *Proceedings of the eCoLoRe-MeLLANGE Workshop,* 39-44.

Selinker, L. 1972. Interlanguage. *International Review of Applied Linguistics* 10(3), 209-231.

Swales, J. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.

Tausczik, Y. and J. W. Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29(1), 24-54.

Toury, G. 1979. Interlanguage and its manifestations in translation. *Meta: Translations' Journal* 24(2), 223-231.

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

Toury, G. 2004. Probabilistic explanations in translation studies: Welcome as they are, would they qualify as universals? In A. Mauranen and P. Kujamäki, eds., *Translation Universals: Do They Exist?*, 15-32. Amsterdam: John Benjamins.

Uzawa, K. 1996. Second language learners' processes of L1 writing, L2 writing, and translation from L1 into L2. *Journal of Second Language Writing* 5, 271-294.

Volansky, V., N. Ordan and S. Wintner. 2015. On the features of translationese. *Digit. Scholarsh. Humanit* 30, 98-118.

Xiao, R. 2009. How different is translated Chinese from native Chinese? *International Journal of Corpus Linguistics* 15, 5-35.

Xiao, R. and G. 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.

Xiao, R. 2015. Contrastive corpus linguistics: Cross-linguistic contrast of English and Chinese. In B. Zou, S. Smith and M. Hoey, eds., *Corpus Linguistics in the Chinese Context*, 98-113. Basingstoke: Palgrave Macmillan.

Wang, W. and Q. 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.

Examples in: English
Applicable Languages: English
Applicable Level: Tertiary

