

# A Corpus-Based Study on Original English Abstracts and Translated English Abstracts: A Case Study of Passive Voice and Pronouns

Mingyao Chen<sup>1</sup> & Qiongxia Ye<sup>1</sup>

<sup>1</sup> School of Foreign Languages, Zhejiang Gongshang University, China

Correspondence: Mingyao Chen, School of Foreign Languages, Zhejiang Gongshang University, Hangzhou, Zhejiang 310018, China. Tel: 86-138-5818-8329. E-mail: chenmingyao@hotmail.com

Received: February 16, 2014 Accepted: March 31, 2014 Online Published: November 25, 2014

doi:10.5539/ijel.v4n6p52

URL: <http://dx.doi.org/10.5539/ijel.v4n6p52>

## Abstract

On the basis of a large amount of corpus-based studies on translation works, the translation universals hypothesis is proposed. As it claims, translations enjoy some general features and Baker (1993) summarizes them into three universals, namely simplification, explicitation, and normalization, which are supported by many following researches.

However, some of the later studies contradict with these rules in several ways, and the usages of passive voice and pronouns are the two most controversial issues. Previous researches suggest that according to the universal features of explicitation and normalization, translated texts tend to have a lower frequency of pronouns while over-representing the passive voice. To examine such claimings, 160 original English abstracts from two leading journals in the field of translation studies, *The Translator and Translation Studies*, and another 160 English abstracts from *Chinese Translator Journal* and *Chinese Science & Technology Translators Journal*, which are translated from Chinese abstracts, are collected. Two corpora are then constructed, namely the Original English Abstracts Corpus (OEAC) and Translated English Abstracts Corpus (TEAC). The CLAWS Part-of-speech Tagger is used to tag the lexical items and word processing tool AntConc 3.2.4 is used for retrieving the words.

The comparison between the two corpora suggests that the translated English abstracts contain a lower level of frequency in the use of both passive voice and pronouns, which partially query the hypothesis of explicitation and normalization. A detailed analysis shows a higher frequency of past-tense passives in the OEAC and more passives in perfect tense in the TEAC. The OEAC also contains more relative pronouns while the other contains more indefinite pronouns. The norm theory is utilized to account for such phenomena. The detailed results of the study are expected to shed some lights on professional translating and academic writing.

**Keywords:** corpus, OEAC, TEAC, passive voice, pronouns

## 1. Introduction

### 1.1 Background of the Study

As the world develops, the connections between different countries become closer, not only in economic relations but also in social and cultural communications. As a result, the important role of translation is realized by the public society, which boosts the research of translation among the linguists. The rapid development of computer-aided corpus linguistics in 1990s on the other side sheds a new light into the studies of translation, which makes it possible for the studies of translations on a large scale. As more and more translation works have been studied on the basis of corpus, some universal features have been found, which are attached to translation works regardless of the translators and the source languages. Thus the translation universals hypothesis has been put forward, first summarized by the British linguist Mona Baker (1993), who states that all the translated works might share a fixed group of characteristic features “as a mediated communicative event”, no matter who the translator is and from which language it was translated.

In the long run, translation studies move from intuitive hypotheses to scientific conclusions, from perceptual understandings to rational judgments, from individual phenomena to general rules, and during this process scientific tests and verifications are indispensable. The emerging of translation corpora in this respect covers the shortage of previous studies, providing a new scientific tool for many practical problems in studying translation issues, expanding the research scope and bringing in a new perspective of study.

### 1.2 Main Purpose of the Study

The main purpose of this study is to examine the use of passive voice and pronouns in the original English abstracts and translated English abstracts to see whether there is any linguistic difference or whether there are any shared features in translated English academic writing. The results will show some clues of whether the translation universal hypothesis is applicable in certain genres and in the texts translated from some certain languages: in this case the academic abstracts translated from Chinese.

### 1.3 Research Objects

#### 1.3.1 Passive Voice

The large number of passive voice is one of the distinctive features of English language, which is used when the doer of an action is unknown, or unimportant, or when the emphasis is “on the experiment or process being described” (Hacker, 2003, p. 130). It is used to present an objective view, which is typically a common phenomenon in English academic writing.

The passives can appear in many different forms. The most basic passive pattern is the short dynamic be-passive in the “be-verb + Past Participles” constructions (Biber et al., 2009, p. 938), while sometimes the “be-verb” is replaced by “become” or “get”, which is extremely rare in academic writings (Biber et al., 2009, p. 476). Sometimes the passive voice is also realized in the form of “have/make + sth. + Past Participles”. The variation of the form of passive voice makes it very difficult for accurate identification with any software on its own. The “Be-verb + Past Participles” construction, which is the most basic and common form of the passives, is what we are mainly investigating in this paper. One common linguistic feature of these constructions is that they all include a “be-verb”, which can be easily recognized and retrieved by computer tools.

#### 1.3.2 Pronouns

Pronouns refer to those items which take the place of a noun. They function as nouns and are used to avoid lexical repetitions, referring to something that has been mentioned before. In addition, pronouns are used where the reference is unknown or very general, and for specific clause-binding functions (Biber et al., 2009, p. 327). Pronouns can be in three cases: subject, object, or possessive, and they can be classified into several subtypes: personal pronouns; reflexive pronouns; possessive pronouns; demonstrative pronouns; interrogative pronouns; relative pronouns; indefinite pronouns. (Quirk et al., 2002, pp. 335-371)

## 2. Literature Review

### 2.1 Translation Universal and Translationese

Translated works used to be regarded as second-hand texts, or poor copies of the original writings with no point to be studied on its own. It became worth investigating “as a system in its own right” only since the proposal of polysystem theory in the late seventies of last century (Baker, 1993). The theory was proposed by Even-Zohar, which sees the literature world as a polysystem, “a hierarchical and dynamic conglomerate of systems rather than a disparate and static collection of texts” (ibid). The translated literature “would not be disconnected from original literature” (Even-Zohar, 1979, qtd. in Baker, 1993).

As more and more corpus-aided translation studies had been conducted, Mona Baker noticed that some linguistic patterns are specific to translated texts, which are not the result of interference from the source or target language, and “inexplicable in terms of any of the repertoires involved” (Even-Zohar, 1979, qtd. in Baker, 1993). Mona Baker (1993) concluded these patterns as “universal features of translation” in her later discussions. The hypothesis of translation universals was then put forward, claiming that any translated language variety might share some groups of fixed characteristic features.

The first universal, simplification, refers to the assumption that the language in translation is assumed to be simpler than that of the target language original, in terms of the lexical usage and syntactic structures (Baker, 1993). The second, explicitation hypothesis, indicates that the translations are likely to be more explicit than the original target language texts (Baker, 1993). The third, normalization, means the exaggeration of some typical features of the target language in translated works, which sometimes is also referred as conventionality, conservatism, or standardization (Toury, 1995, qtd. in Puurtinen, 2003). According to Baker, normalization of the language can be manifested in grammaticality, typical punctuation, and collocational patterns (Baker, 1996).

Besides Mona Baker, some other linguists have also noticed that the translated language contains some distinctive features, which is even seen as a “dialect” within a language. It is called “third code” by Frawley, or “translationese” by Gellerstam (Frawley, 1984; Baroni & Bernardini, 2006).

As translation works gradually gained its status as an independent research subject, the term translationese, used

in a neutral sense, was also put forward, referring to simply the translation-specific language. It was originally described by Gellerstam in 1986 as “the set of ‘fingerprints’ that one language leaves on another when a text is translated between the two” (Baroni & Bernardini, 2006).

### 2.1.1 Supportive Voices

The findings of those studies that conform to the hypothesis include: translators are more careful and conservative in their use of language, so that they prefer to use more standard forms of the language; the translations tend to have a higher level of formality and they are “sanitized”, that is, the translators usually would like to avoid certain features of the language use, such as some irregular spellings and regionalisms; the translators are inclined to produce more “uniform” texts (Baker, 2004). The hypotheses to be introduced in this paper are the universal features summarized by Mona Baker, which include the three characteristics, i.e., simplification, explicitation, normalization.

A pioneering project concerning the translation universals were conducted by Anna Mauranen in Finland in 1998. Linguists involved in this project study the translation universals hypothesis on the basis of corpora analysis in a variety of languages, and then relate their findings with the hypothesis proposed by Mona Baker and try to analyze in the light of language structure (Puurtinen, 2003).

The translation universal hypothesis also receives supportive voice from Gellerstam, who compared the translated and original Swedish novels in 1996, and then found the differences in the use of reporting clauses. (Gellerstam, 1996, qtd. in Baroni & Bernardini, 2006).

Later studies on this issue can also be found in China. Huang investigates the correspondence between the translation and the source texts on the basis of his self-created English-Chinese parallel corpora in 2003. He (2003) finds that the translations always show the sign of amplification, and the degree may vary according to different types of text. This finding can be related to the hypothesis of explicitation, which partly explains this phenomenon.

The study by Wang and Hu in 2008 further proves the three universal features in translational texts. They conclude that the simplification of content words, the explicitation of function words and pronouns as two general features of translated Chinese language.

### 2.1.2 Questionings

There are many other researches on this issue going contradictory with the translation universal hypothesis. Jantunen’s findings in 2001 on synonymous amplifiers in Finnish translations, for instance, do not support the simplification assumption. Puurtinen’s (2003) findings on the use of connectives in Finnish children’s literature partly contradict the explicitation assumption.

## 2.2 Empirical Research on Abstracts

There have been many related studies regarding the features of abstracts. However, most of them are conducted from the perspectives of stylistics, discourse analysis or pragmatics, or in terms of the translating progress, and most of the results are explained in the light of “norm”.

The studies on abstracts can be roughly divided into two types, and the first one concerns translating the Chinese abstracts into English. For example, Liu (2001) introduces the main syntactic structures used in abstracts and specific ways of their translation. The translation of the passive voice and active voice are also elaborated. The other type of researches examines the abstracts themselves, in terms of the language use, such as the tense, voice, personal pronouns. Studies show that international abstracts use more active voice while the Chinese scholars tend to use more passive voice in their abstracts (He, 2004; Fan, 2005). And the international abstracts use “we” more often than the Chinese writers do (Zhang, 2006).

## 2.3 Debates on Passive Voice and Pronouns

### 2.3.1 Debates on the Use of Passive Voice

As one of the most important features of English language, the use of passive voice is also one of the biggest differences between English and Chinese (the source language of the translated texts in this study). According to the translation universal hypothesis of normalization, translators tend to over-represent some typical features of the target language, thus they are likely to use more normalized or standardized expressions.

However, previous studies show some interesting results. The study on original Chinese texts and the ones translated mainly from English, conducted by Zhonghua Xiao and Guangrong Dai, shows that influence by the source language, the translated Chinese texts show a preference for the passive voice which is not so common

used in Chinese original texts (Xiao & Dai, 2010). In this case, the normalization hypothesis of the translated language is not proved while the influence of source language seems to be the dominant features of the translation.

According to the normalization universal, the translations would probably exaggerate some specific features of the target language as translators tend to avoid the risk of making their translations sound exotic and unreadable, so the language in translations is likely to be more conventional. If this hypothesis is related to the present study, a deduction can be made that the passive voice is more highlighted in the translated English abstract than in the original English abstracts.

Therefore, two questions have been raised: are there any obvious differences between the translated English academic abstracts and original ones in terms of the use of passive voice? And if the differences exist, how do they differ from each other? These two questions will be answered and the assumption of a higher frequency of passive voice in translated English abstracts will be verified or falsified in this study.

### 2.3.2 Debates on the Use of Pronouns

The use of pronouns has also been attracting the attention of many linguists and the findings of their researches point to two different sides. According to Laviosa-Braithwaite, “translators may tend to repeat redundant grammatical items, such as prepositions, and overuse lexical repetition, which in turn results in a lower frequency of pronouns” (Laviosa-Braithwaite, 1996). This claim illustrates the explicitation hypothesis. Vanderauwera and Blum-Kulka also have similar findings in their research (Blum-Kulka, 1986, qtd. in Puuriten 2003).

However, there are also some different voices. In Borin and Prutz’s syntax-focused research in 2001, it is found that the translated language has the sign of the over-representation of adverbs, infinitives and pronouns as well (Baroni & Bernardini, 2006). In Wang and Hu’s study on translated Chinese language in 2008, it is found that the higher frequency of function words and pronouns is one of the general features in translated texts (Wang & Hu, 2008).

## 3. Research Method

### 3.1 Corpora Used for This Study

#### 3.1.1 Original English Abstracts Corpus (OEAC)

The first one—Original English Abstracts Corpus (OEAC) consists of 160 academic abstracts (totally 24,046 running words) taken from the two best journals in the field of translation studies—*The Translator*, and *Translation Studies*. *The Translator*, published by St. Jerome Publishing, is seen as the best journal in the field. It is a refereed international journal that publishes articles around a variety of issues related to translation and interpreting and it is also listed in the Arts and Humanities Citation Index and the Social Science Citation Index. Another journal, *Translation Studies*, published by Routledge, with three issues per year, is also among the best journals in translation studies. The journal is abstracted/indexed in Annotated Bibliography of English Studies (Routledge ABES), Bibliography of Translation Studies, Linguistic Abstracts Online and so forth. The articles in these two journals and their abstracts are regarded as having high quality.

There are in total 81 abstracts in *The Translators* and 83 in *Translation Studies* between 2008 to 2013, and in order to make it easier for the study, we only keep 80 abstracts from each journal. We need to stress that the deletions of some abstracts are made indeliberately and the purpose is to keep the balance of the data in the corpus as well as to make it easier for the treatment of numbers.

#### 3.1.2 Translated English Abstracts Corpus (TEAC)

For the second corpus—the Translated English Abstracts Corpus (TEAC) in this study, we collected 160 translated English abstracts from the two leading core translation journals in China: *Chinese Translators Journal*, and *Chinese Science & Technology Translators Journal*.

*The Chinese Translators Journal*, founded in 1979, is hosted by Chinese Foreign Languages Bureau and Translators Association of China (TAC for short). The TAC is the only nation-wide association in the field of translation, and it functions both as an academic society and a trade association. Its bimonthly journal, *Chinese Translator Journal*, is recognized as the best, most authorized and most influential journal in the translation area in China. *The Chinese Science & Technology Translators Journal*, held by the Science & Technology Translators Association of the Chinese Academy of Science (CAS), is another leading journal in the field, which gathers high-quality articles by best translation scholars.

In the present study, 80 translated English abstracts are taken from each journal and the small comparative

corpus is then built as opposite to the OEAC. The average length of the abstracts in Chinese Science & Technology Translators Journal is much shorter than that of the abstracts from *Chinese Translators Journal*, and in general the translated English abstracts are shorter than the original abstracts from the two English core journals. This fact leads to the imbalance between the two corpora in their sizes, and how we deal with this imbalance will be introduced in 3.4.

### 3.2 Research Tools

#### 3.2.1 CLAWS Part-of-Speech Tagger

Part-of-speech (POS) tagging, also called grammatical tagging, is the first step of dealing with corpora. CLAWS (the Constituent Likelihood Automatic Word-tagging System), which was developed by UCREL at Lancaster, has been continuously developed since the early 1980s (Garside & Smith, 1997). As one of the most widely used tagging tool, CLAWS has consistently achieved 96-97% accuracy with an error-rate of only 1.5% within the BNC (British National Corpus) (<http://ucrel.lancs.ac.uk/claws/>, June 15, 2013). The free trial of CLAWS is available at <http://ucrel.lancs.ac.uk/claws/trial.html>.

There have been several tagsets in CLAWS over the years. On the free online website, tagset C5 and C7 (also known as BNC Enriched Tagset) are available. C5 tagset has just over 60 tags while C7 is much more precise with over 160 tags (<http://ucrel.lancs.ac.uk/claws/>, June 18, 2013). For example, tag “VVN” in C7 tagset stands for the past participle form of the lexical verbs, which are essential in English passive voice; “Be-verbs” are tagged as VB0, VBDR, VBG, VBI, VBN etc. according to their different forms. In this study, we adopt the C7 tagset.

#### 3.2.2 AntConc 3.2.4

The software AntConc 3.2.4, used in the present study, is one the most widely used corpus-based language analyzing tools. It was developed by Dr. Laurence Anthony, a professor at Waseda University, Japan (<http://www.antlab.sci.waseda.ac.jp/resume.html>, 2013-06-15). AntConc is a freeware concordancer and corpus analysis tool, which includes a range of functions, such as concordance, clusters, collocates, word list, keyword list, N-grams. There are many different versions of the software, such as AntConc 3.2.4 used in this article, AntConc 3.1.2, AntConc 3.2.0, AntConc 3.3.0, AntConc 3.3.5 etc, with some small differences for instance in the users' interface. Taking into consideration of various factors, we finally choose the AntConc 3.2.4 as the best version. This software is available on Dr. Laurence Anthony's own website [http://www.antlab.sci.waseda.ac.jp/antconc\\_index.html](http://www.antlab.sci.waseda.ac.jp/antconc_index.html), and it is updated for free on a regular term.

The credibility of the corpus tool AntConc has been proved by previous researches (Wang, 2009; Wei et al., 2005). The user's interface of AntConc3.2.4 is very user-friendly, and “.txt” files including lists of words can be loaded via the advanced options for a more convenient searching. The overall distribution of the researched items is displayed under “Concordance Plot” and the detailed contexts of each retrieved word can be seen via “File Views”.

### 3.3 Study Design

This research adopts a hybrid methodology to investigate the linguistic features identified in the academic abstracts on translation studies in original English corpus and translated English abstract corpus. First, the researcher built the Original English Abstracts Corpus and a comparative corpus of translated English abstracts. Second, a comparative study is conducted between the two corpora, regarding the linguistic features of the language, in terms of the use of passive voice and pronouns in each corpus. The results of the comparison will be summarized and analyzed in the light of translation universal hypothesis: whether they support or question the previous findings.

## 4. Research Results

### 4.1 An Initial Comparison of the Two Corpora

An initial comparison between the two corpora shows that there is a huge difference in the total numbers of the words. The one made up of 160 original abstracts (OEAC) has 25,480 words with 171,642 characters (include the space), and the other one TEAC has only 13,869 words with 92,660 characters (space included), as shown in the table below:

Table 1. Initial comparison of the two corpora

	Number of Texts	Total Words	Number of Characters (excl)	Number of Rows	Average Length of Each Abstract (words)
OEAC	160	25,480	146,295	2,038	159.25
TEAC	160	13,869	79,138	1,194	86.68

As shown in the table, the two corpora has a huge difference in their sizes, so in order to make the research more scientific and the results more convincing, and for the ease of comparing, the method of ratio will be used to normalize the results in dealing with the two sets of figures in the following parts of this section.

#### 4.2 Frequency of Passive Voice

As is explained in the first section, the majority of passive voices are realized in the form of “Be-verbs + past participles”, which is what we mainly research in this paper. Then the first thing is to find out all the be-verbs in the two corpora, and then select out the ones needed in this study. As the corpora have been tagged by CLAWS, and according to the tagset C7 system, the “Be-verbs” are tagged as VB0, VBDZ, VBG, etc. according to their different forms. The whole category can be included as “VB\*” (in which “\*” serves as a wildcard) in AntConc 3.2.4 and then retrieved. Then we use the “Advanced Search” to limit the context words of these be-verbs that have been retrieved, and select out those go with past participle form of the lexical verbs, which are tagged as VVN in the text. For further analysis, we also search for the eight forms of be-verb separately, i.e., “is”, “was”, “are”, “were”, “am”, “be”, “being”, “been”, and then delete those irrelevant to passive voice in every retrieval result.. The “get/become + past participles” constructions which signal passive voice are retrieved as well, though it only appears twice in the two corpora. So is the “make/have + sth. + past participles” combination. Besides, there are some verbs that are tagged as VVD (past tense form) which however also indicate a passive voice, and these sentences need to be selected manually. For example, the verb “produced” is tagged as a VVD in the sentence “It then examines a translation of Othello produced by...”, and this sentence is definitely a passive voice. There are also other examples, such as “The study reported here is...”, “the analysis offered here...”.

After the machine retrieving and manual selection, an initial result is displayed in the following.

##### 4.2.1 An Overview

First we will look at the overall distributions of the passive voice in the two corpora. The comparison between TEAC and OEAC shows some slight differences between the translated and original languages as shown in Table 2.

Table 2. Frequency of passive voice

	No. of Words	No. of Passive Voice	Frequency (%)
OEAC	25,480	326	1.28
TEAC	13,869	157	1.13

And the frequency of “Be-verb” indicating a passive voice in the two corpora are shown in Table 3:

Table 3. Frequency of “Be-verb”

	No. of Words	No. of “Be-verb”	No. of “be-verb” Indicating A Passive Voice	Frequency (%)
OEAC	25,480	539	312	1.22
TEAC	13,869	338	150	1.08

From the two tables we can see that 312 expressions of passive voice in the OEAC are realized with a “Be-verb”, sharing 95.71% of all the passives, while in TEAC the percentage is 95.54%. Table 2 and 3 also indicate that the OEAC enjoys a higher frequency of passive voice than TEAC in overall, and so is the frequency of passives with a “Be-verb”.

##### 4.2.2 “Be-Verbs” in Different Tenses

As more than 95 percent of the passive voices in the two corpora are realized by “Be-verb”, in this paper we

mainly compare the use of “Be-verb” as a signal of passive voice in the OEAC and TEAC. The various forms of the be-verb are investigated separately (eight in total), and the results are shown in separate tables in the following part. The be-verb in the form of “am” only exists once in the TEAC, and it is not served for passive voice in the context. If the word is traced back to the context, it can be found that the “am” in the concordance result is a mistake of AntConc due to the failure of recognition of the word “flam”. Thus there is no use of “am” as part of passive voice in both corpora.

The ratio shows the percentage of the passive voice presented with an “is” among all the places of passive voice with a be-verb. First, the use of “is” is the most widely used in the corpora:

Table 4. Frequency of “is”

	No. of Passive Voice	No. of “is”	No. of “is” in Passive Voice	Ratio (%)
OEAC	312	223	97	31.09
TEAC	150	172	50	33.33

Note. The second row “No. of Passive Voice” refers to the number of passive voice with a *be*-verb. And “Ratio” refers to the percent of passive voice presented with an “is” among all the passive voices with a *be*-verb. So are the rest tables in this section.

Table 5. Frequency of “are”

	No. of Passive Voice	No. of “are”	No. of “are” in Passive Voice	Ratio (%)
OEAC	312	84	50	16.03
TEAC	150	65	26	17.33

Seen from the two tables above, there is no big difference in the frequency of “is” and “are” between the two corpora. However when we look at the past form of the “be verb”, some remarkable differences appear.

Table 6. Frequency of “was”

	No. of Passive Voice	No. of “was”	No. of “was” in Passive Voice	Ratio (%)
OEAC	312	66	37	11.86
TEAC	150	4	2	1.33

Table 7. Frequency of “were”

	No. of Passive Voice	No. of “were”	No. of “were” in Passive Voice	Ratio (%)
OEAC	312	35	22	7.05
TEAC	150	3	2	1.33

Table 6 and Table 7 show the significant difference in the frequency of the use of “was” and “were” as indicators of passive voice between the OEAC and TEAC.

The original form of be-verb—“be” is then examined, and the result is displayed as below, followed by the result of the word “being” in Table 9:

Table 8. Frequency of “be”

	No. of Passive Voice	No. of “be”	No. of “be” in Passive Voice	Ratio (%)
OEAC	312	74	55	17.63
TEAC	150	59	42	28

Table 9. Frequency of “being”

	No. of Passive Voice	No. of “being”	No. of “being” in Passive Voice	Ratio (%)
OEAC	312	13	7	2.24
TEAC	150	3	2	1.33

At last we checked the use of “have/has/had+ been + Verb” constructions by searching the word “been”, which

are also common used especially in academic English writings to indicate a passive voice.

Table 10. Frequency of “been”

	No. of Passive Voice	No. of “been”	No. of “been” in Passive Voice	Ratio (%)
OEAC	312	42	25	8.01
TEAC	150	32	26	17.33

#### 4.3 Frequency of Pronouns

As the texts in the two corpora are tagged and ready to be analyzed, we also create a list of pronouns which is then loaded into the AntConc 3.2.4 for advanced search of the concordance in the two files (one translated and one original English abstracts). The context words are sometimes used with manual selection to exclude those irrelevant hits of the item.

##### 4.3.1 An Overview

First we search the frequencies of the whole category of pronouns in these two files, and the result is as shown in the table:

Table 11. A rough comparison in the use of pronouns

	File Length (chars)	No. of Words	No. of Hits	Frequency (%)
OEAC	172,070	25,480	1,574	6.18
TEAC	93,096	13,869	831	5.99

A clearer view of the how these pronouns distribute can be seen by the screenshot of the AntConc as below:

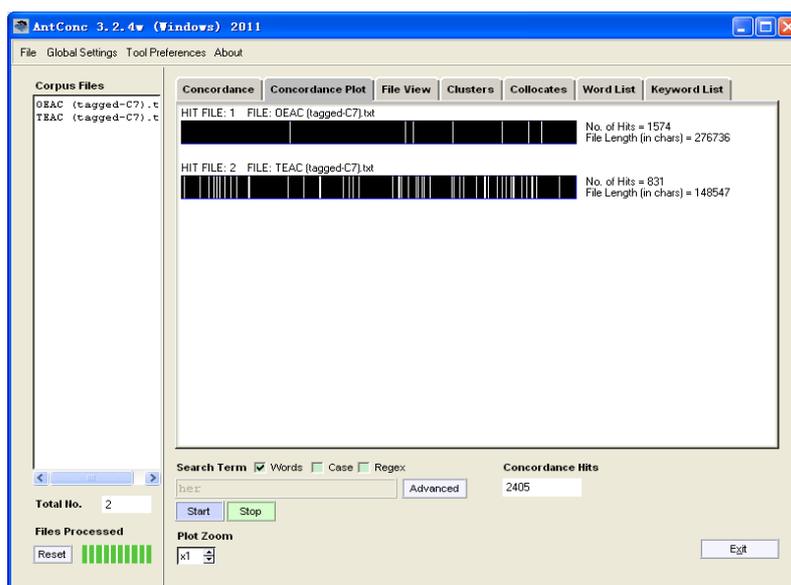


Figure 1. Screenshot of the frequency of pronouns

The FILE 1 in the picture named “OEAC (tagged-C7)” refers to the 160 original English abstracts from *The Translator* and *Translation Studies* which are tagged in CLAWS C7 tagset, and the FILE 2 “TEAC (tagged-C7)” refers to the 160 tagged English abstracts translated from Chinese. The figures in the screen shot only shows the initial result generated by the software, which is not reliable on its own as further manual selection is required to exclude those irrelevant uses of “that” and “one”.

What else, some pronouns belong to more than one subtypes. For example the word “that” belongs to demonstrative pronouns and relative pronouns, and many of those interrogative pronouns also belong to relative pronouns, such as “which”, “who”, “whom”, and “whose”. Therefore machine retrieving is combined with

manual selection to revise the results. A closer and more detailed view of the frequency of each group of pronouns will be examined in the coming sections.

#### 4.3.2 A Closer View on Each Subtype

With an overview of the frequency of pronouns in the two corpora, we then further investigate the use of each subtype of pronouns in these files to see what will happen.

The Personal pronouns (I, me, we, us, you, you, she, her, he, him, it, they, them) are firstly checked, which are tagged as PPH1 (singular personal pronouns), PP02 (plural personal pronouns), PPHS1 (singular personal pronouns, third person) etc. The noticeable fact is that the word “her” belongs to both personal pronouns and possessive pronouns, and there are only two places where “her” is used as personal pronouns in the two corpora, both of which exist in the OEAC, which are tagged as PPPH01 and the other 45 places of “her” used as possessive pronouns are tagged as APPGE. After manual selection, the final result is then achieved which is encouraging:

Table 12. Frequency of personal pronouns

	File Length (in chars)	No. of Words	No. of Hits (personal pron.)	Frequency(%)
OEAC	172,070	25,480	320	1.26
TEAC	93,096	13,869	153	1.10

*Note.* The frequency in this table refers to the ratios of personal pronouns among the whole word counts in the two corpora, so are other tables in 4.3 and 4.4.

Then comes the reflexive pronouns (myself, yourself, himself, herself, itself, ourselves, yourselves, and themselves).

Table 13. Frequency of reflexive pronouns

	File Length (in chars)	No. of Words	No. of Hits (reflexive pron.)	Frequency (%)
OEAC	172,070	25,480	30	0.12
TEAC	93,096	13,869	13	0.09

The investigation on the frequency of possessive pronouns (my, your, his, her, its, our, your, their, mine, yours, his, hers, ours, yours, theirs) (tagged as PNX1, PPX1, PPX2 etc.) also involves in the manual selection of the word “her”, and the result shows something interesting which indicates an opposite tendency to the previous groups of data:

Table 14. Frequency of possessive pronouns

	File Length (in chars)	No. of Words	No. of Hits (possessive pron.)	Frequency (%)
OEAC	172,070	25,480	249	0.98
TEAC	93,096	13,869	167	1.20

Different from other two subtypes of pronouns examined above, the TEAC shows an obviously higher level of frequency of possessive pronouns than the OEAC.

The result of demonstrative pronouns (this, these, that, those) generated by the software is shown in Table 15 in the following. In line with the overall frequency of pronouns, the TEAC uses much less demonstrative pronouns than the OEAC:

Table 15. Frequency of demonstrative pronouns

	File Length (in chars)	No. of Words	No. of Hits (demonstrative pron.)	Frequency (%)
OEAC	172,070	25,480	629	2.47
TEAC	93,096	13,869	308	2.22

However, the very important point to be noticed is that the word “that” belongs to both demonstrative pronoun and relative pronouns which should be examined further. In addition, it can also be used as a conjunction “after certain verbs, nouns and adjectives to introduce a new part of the sentence” (Wehmeier, 1993). For example, the word “that” in the sentence “I don’t know that he will also come”. Besides, “that” has another role as an adverb, similar to “so” and “such”, when it is used to describe the degree or extent (ibid). For instance, “At the moment I can’t walk that far.” What’s more, as the retrieving is done by the machine, the word “that” in some phrases such as “so that” cannot be excluded as well. These three uses of “that” are irrelevant to the present part of study, so the result of the automatic retrieval should be examined manually in the coming section, and a new comparison will be made after the selection and classification of “that”.

Then comes the comparison of the use of the interrogative pronouns (what, who, which, whom, whose). Overlapped with relative pronouns, the frequency of interrogative pronouns is very low in the academic writings, and the two corpora enjoy almost the same frequency of this subtype of pronouns.

Table 16. Frequency of interrogative pronouns

	File Length (in chars)	No. of Words	No. of Hits (interrogative pron.)	Frequency (%)
OEAC	172,070	25,480	15	0.06
TEAC	93,096	13,869	8	0.06

However, the difference on the frequency of relative pronouns (that, which, when, what, who, whom, whose, whichever, whoever, whomever) between the two corpora is quite obvious:

Table 17. Frequency of relative pronouns

	File Length (in chars)	No. of Words	No. of Hits (relative pron.)	Frequency (%)
OEAC	172,070	25,480	436	1.71
TEAC	93,096	13,869	167	1.20

As is explained above, the results in this table also include the use of “that” which should be examined again and selected in the next section. This table only provides some clues for a vague idea and cannot be wholly reliable.

The initial comparative results of the frequency of indefinite pronouns generated by automatic calculation of the AntConc software are shown below:

Table 18. Frequency of indefinite pronouns

	File Length (in chars)	No. of Words	No. of Hits (indefinite pron.)	Frequency (%)
OEAC	172,070	25,480	201	0.79
TEAC	93,096	13,869	144	1.04

The word “one” in this search result should also be examined further, as it can be used as a quantifier in addition to a pronoun. Further investigation should also be conducted in the next section to exclude those irrelevant uses of these words in the two corpora, or any possible mistakes made by the CLAWS and AntConc.

#### 4.3.3 Manual Selection of “That”

As explained in the previous section, the use of “that” in the two corpora should be examined further with manual selection, deleting those ones that do not function as pronouns. If we investigate the frequency of “that”

on its own, the result is as shown in the third column of Table 19. And then all these uses of “that” are retrieved back to the context to differentiate whether they are used as pronouns or something else. A further classification between the use of “that” as demonstrative pronouns and relative pronouns is also made by hand for a clearer result. As we can see from the results, the OEAC has a much higher frequency of the use of “that” either as a demonstrative pronoun or as a relative pronoun.

Table 19. Different use of “that”

	No. of Words	No. of Hits (“that”)	No. of “that” Used as Demonstrative Pron.	No. of “that” Used as Relative Pron.
OEAC	25,480	284 (1.11%)	22 (0.09%)	134 (0.53%)
TEAC	13,869	107 (0.77%)	6 (0.04%)	21 (0.15%)

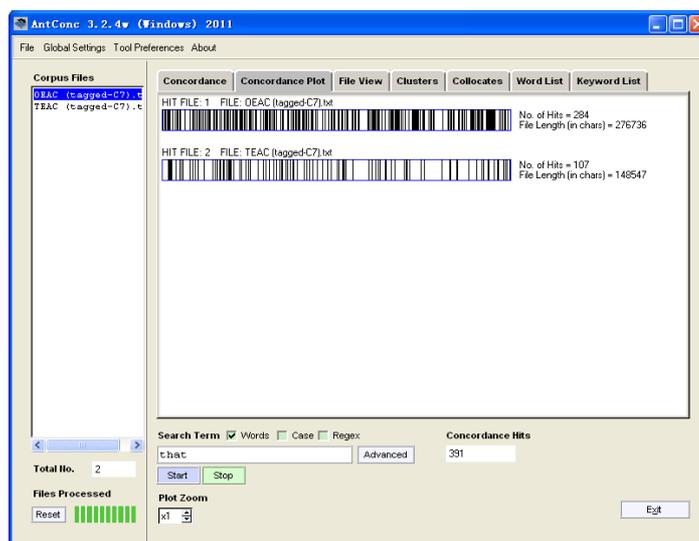


Figure 2. Screenshot of the frequency of “that”

The figure above shows the screen shot of the result generated by the software.

If we relate this result to the previous automatic ones generated by the software in the last section, as shown in Table 15 and Table 17, then two new groups of data of the frequency of demonstrative pronouns and relative pronouns in the two corpora can be worked out:

Table 20. Frequency of demonstrative pronouns II

	No. of Words	No. of Demonstrative Pron. Calculated by AntConc	No. After Manual Selection	Frequency (%)
OEAC	25,480	629	367	1.44
TEAC	13,869	308	207	1.49

Table 21. Frequency of relative pronouns II

	No. of Words	No. of Relative Pron. Calculated by AntConc	No. After Manual Selection	Frequency (%)
OEAC	25,480	436	286	1.12
TEAC	13,869	167	81	0.58

As we can see from the results, the difference in the frequency of demonstrative pronouns between the two corpora is not so obvious, however the difference in the frequency of relative pronouns is quite noticeable: the OEAC has a much higher frequency of relative pronouns than TEAC, with the former being almost twice of the latter.

#### 4.3.4 Reinvestigation of “One”

Similar to “that”, the use of the word “one” is also reinvestigated on its own, and it is then categorized according to its different roles in the sentence as an indefinite pronoun or as a quantifier.

Table 22. Different use of “one”

	No. of Words	No. of Hits (“one”)	No. of “one” Used as A Pron	Frequency (%)
OEAC	25,480	48	15	0.58
TEAC	13,869	24	9	0.65

*Note.* Since the values of this group of figures are too small and the percentage cannot show any difference, the permillage is then used to indicate any minor difference.

There are 9 out of the 24 places in the TEAC where the word “one” serves as a pronoun, such as in the sentences listed below:

- 1) “fluency is in essence a target-reader-oriented concept, *one* which obliges translators to repress...”
- 2) “the translation of Mao’s poetic works was more a political activity than a literary *one*.”
- 3) “Among its extant versions, the earliest *one* is a Ming-dynasty transliteration in Chinese.”
- 4) “inadequate translation of public signs stands out as an especially serious *one*.”
- 5) “Given the interdisciplinary nature of translation and the varied translation theories—*one* must take particular care to...” (TEAC)

And in the other 15 places, the word “one” does not serve as a pronoun, such as in examples below:

- 1) “Drawing from its author’s personal experiences in interpreting, which confirm that a one-hundred-percent reproduction in SI is neither possible nor necessary...”
- 2) “One of the challenges in C-E translation of legal texts is to find English equivalents for words and terms in the Chinese original...”
- 3) “...and many of them are in one way or another caused by the translator’s failure to detect shades of semantic differences in seemingly exact equivalents of the two languages.”
- 4) “Comparing the C-E translations of one text by two groups of students with different language embodiment backgrounds...”
- 5) “Memory enjoys a close and complex relationship with interpreting and is one of the fundamental elements in the understanding of interpretation.” (OEAC)

In OEAC, there are 15 places where “one” is used as a pronoun, and examples are shown below: Due to the limitation of length, we will only list the numbers of the use of “one” as a pronoun in another five files respectively, and the examples of it used as a quantifier will not be displayed.

Examples

- 1) “...as a counter-translation of geographies, namely as the rewriting of a West-oriented, Atlantic geo-poetics into an East-oriented, Mediterranean one.”
- 2) “The first two were done at different junctures in the context of colonialist oppression—one in 1909 when the incipient nationalist movement was in its militant phase...”
- 3) “The role of interpreters and translators in relation to violent conflicts is a complex, dynamic and multi-faceted one, whether...”
- 4) “If translation studies were a country, it would be one that needs a new basis for its domestic policies...”
- 5) “...propose a composite model of analysis of conflicts and their translation, one that grounds itself in specific situations of power and...” (TEAC)

Examples of “one” as a quantifier in OEAC:

- 1) “This paper offers a case-study of one moment in the modern genesis of the term...”
- 2) “Gustave Le Bon (1841-1931) was one of the most important and popular social thinkers of the Third Republic in France.”

- 3) "... is one of an open-ended, network-like constellation of positionings that are..."
- 4) "...destruction and forgetting on the one hand, and gain, survival and remembering of Kurdish culture on the other."
- 5) "...the ways in which translation and code switching may be exploited in the creation of song lyrics featuring more than one language and..." (OEAC)

Then if we delete those examples where "one" does not function as a pronoun from the result generated by the software in the last section, we will get a new group of data which are more reasonable on the frequency of indefinite pronouns examined before. And then we calculate the ratio of the frequency of indefinite pronouns in the two corpora.

Table 23. Frequency of indefinite pronouns II

	No. of Words	No. of Indefinite Pron. Calculated by AntConc	No. After Manual Selection	Frequency (%)
OEAC	25,480	201	168	0.66
TEAC	13,869	144	129	0.93

After the initial computer retrieval and further manual selections, a more reliable group of data of the frequency of pronouns in the two corpora can be reached, as shown in Table 24:

Table 24. Frequency of pronouns after manual selections

File	No. of Words	No. of Pronouns Calculated by AntConc	No. After Manual Selections	frequency (%)
OEAC	25,480	1,574	1,435	5.63
TEAC	13,869	831	758	5.47

## 5. Discussions

### 5.1 Differences in the Use of Passive Voice

The results in section 4.2 show that there are some obvious differences in the use of passive voice between the original English texts and English translations. In overall, the translations have a slightly lower level of the frequency of passive voice. There are totally 326 places of passive voice in the OEAC and 157 in the TEAC, which account for 1.28% and 1.13% respectively. Among them, 312 places of passive voice in the OEAC go with *be*-verbs, while in the TEAC there are 150 places.

This difference is not that noticeable, however, it more or less indicates an opposite tendency to the normalization hypothesis.

#### 5.1.1 Different Preferences of Tenses

If we look at the use of passive voice more carefully, we can find some interesting differences among those different forms. Among all of the passive voices discussed in this study, the percentages of those in the forms of "is/are + verbs" in the two corpora are quite similar, with the ones in TEAC being slightly higher. But the rest forms of passive voice enjoy some big differences. On one hand, The passive voices in the forms of "was", "were", and "being" are used far more often in the original English texts. The biggest difference lies in the use of "was": in OEAC, 11.86% of the passive voices are presented in form of "was + verbs" while in TEAC, the percentage is only 1.33%—only two places among all. There also lies a big difference in the use of "were" to indicate a passive voice—7.05% and 1.33% in OEAC and TEAC respectively. On the other hand, the original English uses less "be" and "been" to realize a passive voice than the translated one does. The percentage of "has/have/had been+ Past Participle" constructions in the original English abstracts are only about half of that in the translated English. The numbers of "be" in the OEAC and TEAC are 55 and 42 respectively, occupying 17.63% and 28% in each corpus.

A clearer view can be seen by the table below:

Table 25. Frequency of each form of “Be-verb” (%)

	“am”	“is”	“are”	“was”	“were”	“being”	“be”	“been”
OEAC	0	31.09	16.03	11.86	7.05	2.24	17.63	8.01
TEAC	0	33.33	17.33	1.33	1.33	1.33	28	17.33

The result of the comparison on the frequencies of passive voice in translated English and original English queries the translation universal of normalization though it partially supports it. Even the same linguistic feature should be viewed differently: some forms of passive voice are used more often in translations while some of them are more frequent in original texts, which indicates that translations and originals have different preferences in the way of using passives.

Roughly speaking, the original English writers have an obvious preference for the past tense and present progressive tense of passive voice and on the other hand, the Chinese translators tend to use slightly more present tense and strikingly more perfect tense which is indicated by the “has/have/had + been + verb” structure.

### 5.1.2 Potential Reasons

This phenomenon may be partly explained by the theory of norms. During the past years, translating is seldom regarded as creative work but a “copy” of the original. Translations have long been put aside in the literature-centered society. People expect a translation work to be a copy of the source text in another language rather than a recreation of his/her own, and the quality of a translation is always regarded to be related to the degree of how proficient the language looks like. Therefore, in doing translation, translators are likely to try their best to produce a translation of high quality in the target languages. In the Chinese education system, the use of “be” and the perfect tense are taught very late during the process of English learning, which are seen as a higher level of English usage than other forms of “be” verbs. Therefore, translators tend to use more of these “senior” forms of passive voice in their English translations, consciously or unconsciously, to show their proficiency of the English language.

## 5.2 Use of Pronouns

The study on the use of pronouns in OEAC and TEAC shows that the translated and original texts differ in their preferences of types of pronouns. In total, the TEAC contains less pronouns than the OEAC does, which conforms to the universal of explicitation in translations, which indicates that the translations tend to elaborate the original texts and they are usually more explicit in the languages, in other words, the explicitation hypothesis suggests that translations tend to use more lexical repetitions instead of using pronouns. Therefore in overall, the result of investigation into pronouns displayed in Table 24 suggests the possible existence of explicitation universal in translations, which tends to lead to an avoidance of pronouns.

### 5.2.1 Different Preferences of Subtypes

Though the overall result shows a probability of the avoidance of pronouns in translated abstracts, detailed investigation suggests the inconsistency among the use of each subtype of pronouns between the two corpora. For example, the personal pronouns in the original texts are used more often than in the translated abstracts, and that the gap is most significant in the use of relative pronouns: the frequency of relative pronouns in OEAC is nearly twice the latter. The difference of reflexive pronouns between the OEAC and TEAC is not so obvious, with the OEAC enjoying a slightly higher frequency than TEAC.

The other three types of pronouns, however, show an opposite tendency. There are far more possessive pronouns and indefinite pronouns in the translated abstracts, and the demonstrative pronouns are also used more frequently in translated texts.

Both of the corpora have almost the same frequency of interrogative pronouns.

The comparisons between the two corpora are as below:

Table 26. Frequency of each group of pronouns

	OEAC (%)	TEAC (%)
Personal Pronouns	1.26	1.10
Reflexive Pronouns	0.12	0.09
Possessive Pronouns	0.98	1.20
Demonstrative Pronouns	1.44	1.49
Interrogative Pronouns	0.06	0.06
Relative Pronouns	1.12	0.58
Indefinite Pronouns	0.66	0.93

One thing to mention here is that the frequency of each form of “Be-verb” in Table 25 refers to the ratio of that form of be-verb among the number of all passives with a be-verb, while in Table 26 the figures mean the frequency of each subtype of pronouns among the whole word count of the corpus. This is due to the different research steps in the process of investigating passive voice and pronouns, and the complexity of further manual selections when dealing with pronouns. In order to keep unified in the way of displaying the results, and for an easier comparison, another table is made which shows the ratio of each subtype of pronouns among the whole number of pronouns.

Table 27. Ratio of each subtype among the whole number of pronouns

	OEAC (%)	TEAC (%)
Personal Pronouns	22.30	20.18
Reflexive Pronouns	2.09	1.72
Possessive Pronouns	17.35	22.03
Demonstrative Pronouns	25.57	27.31
Interrogative Pronouns	1.05	1.06
Relative Pronouns	19.93	10.69
Indefinite Pronouns	11.71	17.02

Though displayed in different ways, the results in these two tables show the same tendencies and the two group of figures in Table 26 and Table 27 are quite conform with each other.

### 5.2.2 Interesting Findings on the Use of “That”

In addition to the different preferences on the use of each subtype of pronouns, the manual selection of the word “that” also brings some interesting findings that are worth noticing.

Seen from Table 19, the OEAC has an obviously higher frequency of the use of “that” either as a demonstrative pronouns or as a relative pronouns: there are in total 107 places of “that” in the TEAC and 284 in the OEAC, among which 6 and 22 are used as demonstrative pronouns respectively; the noticeable fact is that the number of “that” used as a relative pronoun in the TEAC is only 21 while in the original English abstracts there are 134 (the percentages are 0.15% and 0.53% respectively). This huge difference, as well as the results shown in Table 26 and Table 27, indicates that translators and original writers prefer quite different ways in using pronouns.

This result, however, is interesting when compared to the study by Olohan and Baker in 2000. They compare the Translational English Corpus (TEC) and British National Corpus (BNC) in terms of the use of “that”, and find out that the “that-connective” is far more frequent in TEC than in BNC (Olohan & Baker, 2000). Based on this study, Dorothy Kenny also conducted a research on the use of “that” as a optional conjunction in 2004. The research result suggests that translated English is grammatically more explicit than original English, as the optional “that” is far more frequently used in the translated English than the original ones (Kenny, 2004).

The two opposite results may shed some light on further studies, in which we can examine the use of “that” in other genres rather than academic writing, or conduct a study that is based on a larger scale. All in all, this topic is worth further exploration.

### 5.2.3 Possible Reasons

Indefinite pronouns are used to refer to entities which the writer cannot or does not want to specify more exactly (Biber et al., 2009, p. 351). The TEAC writers, or translators, show an obvious preference for indefinite pronouns. This result may attribute to the different writing habits of western scholars and Chinese scholars. Chinese people are accustomed to a much more roundabout way of expressing themselves. This type of

conservative attitude is also reflected in their academic writings.

The possible reason of the strikingly higher frequency of “that” in the OEAC both as demonstrative pronoun and relative pronoun may also be related to the conservative attitude of Chinese writers, and in addition, the replaceable role of “that” in English. When served as a demonstrative pronoun, the reference of “that” is usually vague and it fits in with the use of other vague expressions in conversation (Biber et al., 2009, p. 350). One of the possibilities is that the Chinese writers are afraid of causing any ambiguity in their translations by using “that” as a demonstrative pronoun. When served as relative pronouns, “that” usually can be replaced by other pronouns, such as “which” and “who”, which are regarded as more formal in English teaching system in China. According to norm theory, Chinese writers would prefer to translate their academic abstracts into “better” English versions which are of more formality and proficiency in their eyes. Therefore, they use less “that” as relative pronouns than the original English writers do.

### 5.3 Average Length of Each Abstract

Another implication that can be drawn from the research results displayed in previous sections has nothing to do with linguistic features but cannot be overlooked from the result—that is the average length of each abstract in the two corpora. As introduced in the beginning, each corpus contains 160 abstract from two core journals respectively, however there is a huge difference in the number of words between these two corpora, as shown in Table 27: the number of words in OEAC is almost two times of that in TEAC.

Table 28. A Comparison of the average length

	Number of Texts	Total Words	No. of Characters (excl)	Average Length of Each Abstract (words)
OEAC	160	25,480	146,295	159.25
TEAC	160	13,869	79,138	86.68

This big difference in the length should be mainly due to the different requirements of the English journals and Chinese journals on abstracts. However, if we compare the translated abstracts with their Chinese originals, we will find another potential reason for this phenomenon, that is, translators tend to choose a more efficient translation from several possible versions. For example, translators are likely to use one complicated sentence containing a lot of information instead of several short ones when possible.

## 6. Conclusions

### 6.1 Main Findings and Implications

Based on the research results, the researcher finds that: 1) the original English writers have an obvious preference for the past tense and present progressive tense of passive voice; 2) the Chinese translators tend to use slightly more present tense of passives and strikingly more the “higher-level” perfect tense of passive voice; 3) the original English writers use far more relative pronouns and a little more personal and reflexive pronouns than the Chinese scholars; 4) Chinese translators prefer to use more indefinite pronouns and possessive pronouns; 5) the original English abstracts are much longer than the translated ones on average.

The first aspect of the implications of this study is concerned with the translation universal hypotheses. The research to some degree supports the universals of simplification of translations, and partially prove the explication rules of using less pronouns. The examination on the passive voice in both translated abstracts and original abstracts shows that the English translations of Chinese abstracts still have a lower frequency of this typical linguistic feature of English language, which goes contradictory with the existing normalization assumption. This may be due to the interference of Chinese as the mother tongue which does not feature on passive voice. The overuse of the “been + verb” structure in the translated abstracts can be partly due to the expectancy norms in translations. The results of the whole study suggest that the translations and originals have different preferences in the ways of using the language.

Another implication of the present study lies in its inductive role to those whose work engages in translating. As is indicated by the research findings, original English abstracts would use far more relative pronouns and reflexive pronouns and less possessive pronouns and indefinite pronouns. For the translators, especially those who deal with academic materials, the results of the study may help them to translate into a more native-like text. For the scholars who are interested in this area, this study may serve as some referential data and pilot research which might be inspiring in some way. For those Chinese students and teachers who are working towards the

publishing of their works, this thesis can shed light on their ways of writing abstracts in that they can learn how to write not only abstracts but other academic works as well in a more native way.

### 6.2 Limitations of This Study

#### 1) Asymmetrical corpora

The imbalance in the length of the texts between the TEAC and OEAC does exist. The numbers of the abstracts in the two corpora are the same. However, due to the different requirements of the journals and some other potential reasons, the length of abstracts from different journals varies. This leads to the fact that the word number in the OEAC is almost two times of that in the TEAC. In order to make it convenient to compare the two corpora, the two sets of figures are normalized and presented in percentage.

#### 2) Size of corpora

One of the biggest limitations of the present study is the size of the corpora. Due to various limitations, such as the copyright of these abstracts, the limitation of online downloading, requirement for the quality of the research material, the limitation of time and energy, this research is left on a small scale, serving as a pilot study. Though the normalization is used during the comparison of the corpora which to a large extent increase the reliability of the data, the small amount of the texts involved limits the results of research only to be referential rather than declarative.

#### 3) Accuracy of the software

As this study involves the use of computer software, which to a large degree ease the work while sometimes they are also proved to have some bugs. Take the CLAWS Part-of-speech Tagger used in this research for example: Though authoritative data shows that the accuracy of CLAWS has continuously achieved 96-97% with an error-rate of only 1.5%, there do exist times when the system fails to tag out the grammatical items correctly. This may lead to the existence of a small error of the data generated by AntConc, which, however, should not have big influences on the final result.

### 6.3 Suggestions for Further Studies

The corpora used in this study are not that big in their sizes, therefore the result of this research may not be sufficiently reliable. Further studies can compare the linguistic differences on a larger scale, with the works by native authors of the target language to build the comparative corpus in order to have a better comparison with the translated language.

The different preferences of the forms of passive voice between original and translated language is quite worth investigating in further studies, with a more detailed demonstration and an analysis in depth.

Last but not least, later studies can investigate more materials in academic writings in other subjects. The present paper only examines abstracts in the field of translation studies, and whether the results of the comparison between original English academic writings and translated ones in other subjects, for instance natural science, or engineering, will show similarity or discrepancy with the present research, requires further proof.

## References

- Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, & E. Togini-Bonelli (Eds.), *Text and Technology: In Honor of John Sinclair* (pp. 233-250). Amsterdam & Philadelphia: John Benjamins. <http://dx.doi.org/10.1075/z.64>
- Baker, M. (1996). Corpus-based translation studies: The challenges that lies ahead. In H. Somers (Ed.). *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager* (pp. 175-186). Amsterdam & Philadelphia: John Benjamins. <http://dx.doi.org/10.1075/btl.18.17bak>
- Baker, M. (2004). A Corpus-based View of Similarity and Difference in Translation. *International Journal of Corpus Linguistics*, 9(2), 167-193. <http://dx.doi.org/10.1075/ijcl.9.2.02bak>
- Biber, D., Finegan, E., Johansson, S., Conrad, S., & Leech, G. (2009). *Longman Grammar of Spoken and Written English*. Foreign Language Teaching and Research Press.
- Garside, R., & Smith, N. (1997). A hybrid grammatical tagger: CLAWS 4. In R. Garside, G. Leech, & A. McEnery (Eds.), *Corpus Annotation: Linguistic Information from Computer Text Corpora* (pp. 102-121). London: Longman.
- Hacker, D. (2003). *A Writer's Reference* (5th ed.). Boston: Bedford/St Martin's.
- Kenny, D. L. (2004). *Creativity in Translation: A Corpus-based Study*. St Jerome Publishing Ltd.

- Laurence Anthony's AntConc [EB/OL]. Retrieved from [http://www.antlab.sci.waseda.ac.jp/antconc\\_index.html](http://www.antlab.sci.waseda.ac.jp/antconc_index.html)
- Laurence Anthony's Resume [EB/OL]. Retrieved from <http://www.antlab.sci.waseda.ac.jp/resume.html>
- Laviosa-Braithwaite, S. (1996). *The English Comparable Corpus (ECC): A resource and a methodology for the empirical study of translation* (Unpublished doctoral dissertation). University of Manchester.
- Olohan, M., & Baker, M. (2000) Reporting that in Translated English: Evidence for Subconscious Processes of Explication? *Across Languages and Cultures*, 1(2), 141-158. <http://dx.doi.org/10.1556/Acr.1.2000.2.1>
- Puurtinen, T. (2003). Genre-specific Features of Translationese? Linguistic Differences between Translated and Original Finnish Children's Literature. *Literary and Linguistic Computing*, 18(4), 389-406. <http://dx.doi.org/10.1093/lc/18.4.389>
- Quirk, R. (1985). *A Comprehensive Grammar of the English Language*. London and New York: Longman Group Limited.
- Wang, K. F., & Hu, X. Y. (2008). Corpus-based study of English and Chinese Lexis. *Chinese Translators' Journal*, (6), 16-21.
- Wehmeier, S. (1993). *Oxford Wordpower Dictionary*. Oxford University Press.

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).