NANYANG TECHNOLOGICAL UNIVERSITY

Contrastive Analysis of Pronouns across English, Mandarin Chinese and Japanese

Final Year Project 2013

Name : Yu Jie Seah

Supervisor: Associate Professor Francis Bond

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Associate Professor Francis Bond, for his guidance and support throughout this arduous and challenging FYP period. He gave me lots of valuable help and advice when I felt confused, was understanding when I could not meet deadlines, replied my emails with such amazing speed and was always so pleasant to talk to. Without him, I would not be able to successfully finish this piece of work.

I would also like to thank Research Fellow Wang Shan for her help and guidance during this period. She has always initiated offers of assistance and was very patient in her explanations when I needed them. I appreciate all her comments and opinions to make my FYP a better one.

Lastly, special shout outs to my fellow FYP peers – Sheryl, Tarandip, Delia, Hui Ching, Wei Jia, Hui Ting, Charmaine and more who have gone through this similar experience and have thus emerged as winners! Proud of you guys and of myself, of course! I would also like to thank my family and friends for encouraging me and being there for me when I'm not there for them through this difficult period! Love you guys so very much! ☺

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

A qualitative and quantitative approach was used in this study to examine the distribution of pronouns in three languages, namely English, Mandarin Chinese and Japanese based on the parallel NTU Multilingual Corpus (NTU-MC) with English being the source language while Mandarin Chinese and Japanese translations are aligned to it at the sentence level. The pronouns are extracted from four subcorpora – two short stories, one essay and the other is an online article about Singapore's tourism. However, due to time and space constraints, only pronouns from one subcorpus - *The Adventure of the Speckled Band,* a short story from the Sherlock Holmes series, is tagged, annotated and linked in the corpus. The results show that although English has the most number of pronouns, Mandarin Chinese has the highest percentage of referential pronouns. Also, English has more translated counterparts in Mandarin Chinese as compared to Japanese. We attributed this to the difference in usage of pronouns in the languages. Deprominalisation, surprisingly, was even for both corpora. We believed this to be due to influence from the English text. Findings from this study can shed some light concerning translation issues on pronoun usage for learners of the languages and also contribute to pronoun translation across languages.

CHAPTER ONE

1. INTRODUCTION

Pronouns are an important group of word class in languages. The way they are employed in different languages is interesting to many linguists. Furthermore, in such a globalized world like today, languages are always translated into other languages. Other than translation of content words, how pronouns are translated from language to language can allow one to learn a lot about the language and its translation. English, being the world's most globalized language, has been translated into many different languages. Comparing its translation to Mandarin Chinese and to Japanese can shed light on the usage of pronouns in each language.

1.1 Pronouns in English

Pronouns are a closed class of words (Carter and McCarthy, 2006) that are one of the most commonly seen in the English language (Balogh, 2003). Pronouns are used for their function of replacement of nouns in noun phrases (Carter and McCarthy, 2006). For example:

1a) ***Bobby*** went for a walk.
1b) ***He*** went for a walk.

In the above example (1a) and (1b), we can see that '*he*' can be used to replace '*Bobby*' in the noun phrase. However, if we take out (1a) and only look at (1b), we would not be able to tell whom 'he' refers to. This shows that the context the pronoun is in plays a major role in determining our interpretation of the pronoun (Carter and McCarthy, 2006).

In English, there are several different categories of pronouns. The table below shows the different categories and some examples that belong to each category.

Category	**Examples**
Personal Pronouns	*I, you, he, she*
Possessive Pronouns	*my, his, her, mine, ours, theirs*
Reflexive Pronouns	*myself, herself, himself*
Reciprocal Pronouns	*each other, one another*
Indefinite Pronouns	*anything, somebody, everyone, few, both, neither*
Relative Pronouns	*who, which, what, that, when, where*
Interrogative Pronouns	*who, what, where, when*
Demonstrative Pronouns	*this, that, these, those*

Table 1.1: Examples of pronouns in English

1.2 Pronouns in Mandarin Chinese

In Mandarin Chinese, for personal pronouns, there are three types – first person, second person and third person. Singular and plural forms are shown in the table below (Ross and Ma, 2006). The only reflexive pronoun in Mandarin Chinese is also listed.

	Singular	Plural

First person	我 *wo3* I/me	我们 *wo3men* We/us (exclusive or neutral) 咱们 *zan2men* We/us (inclusive)
Second person	你 *ni3* You 您 *nin2* You (polite)	你们 *ni3men* You
Third person	他 *ta1* He/him 她 *ta1* She/her 它 *ta1* It	他们 *ta1men* They/them (masculine or non-specific for gender) 她们 *ta1men* They/them (feminine) 它们 *ta1men* They/them (non-human or inanimate)
Reflexive	自己 *zi4ji3* Self	

Table 1.2: Summary of personal pronouns in Mandarin Chinese

As we can see from the table above, there are fewer personal pronouns as compared to English. This is due to the fact that Mandarin Chinese is not an inflectional language. Likewise, since Mandarin Chinese does not have case markers (Li and Thompson, 1989), the same personal pronoun is used to represent both the subject and the object.

Also, according to Ross and Ma (2006), there are no possessive pronouns in Mandarin Chinese. Therefore, we express possessives in Mandarin Chinese by adding the particle '的 *de4*' to the pronouns such as '我的 *wo3d*e mine'. However, Ng (2011) has shown that the particle '的 *de4*' can be omitted in Mandarin Chinese when expressing possessives as well, depending on certain factors such as postpositions and the alienability of the possessee nouns.

Similar to English, interrogative pronouns exist in Mandarin Chinese too. They are namely '谁 *shei2* who', '谁的 *shei2de* whose', '什么 *shen2me* what' and '哪儿 / 哪里 *na3er/na3li3* where'. These question words are considered as pronouns due to them being able to head a noun phrase (Li and Thompson, 1989).

Mandarin Chinese also has demonstrative pronouns which are '这 *zhe4* this' and '那 *na4* that' respectively. For plural forms of the demonstrative pronouns, the measure word '些 *xie1*' is added to become '这些 *zhe4xie1* these' and '那些 *na4xie1* those'.

Other pronouns that exist in Mandarin Chinese grammar are the reciprocal pronoun '彼此 *bi3ci3* each other' (Sun, 2006) and indefinite pronouns such as '大家 *da4jia1* everybody' (Yip and Rimmington, 1997).

Being a topic-prominent language (Li and Thompson, 1989), Mandarin Chinese often omits the pronouns after the topic is established. According to Li and Thompson (1989), these omissions are actually zero pronouns where there is an "understood noun phrase referent". For example,

1c) 这　棵　树　∅ 叶 子 很　大。
Zhe4 ke1 shu4 ∅ ye4 zi hen3 da4
'This tree, (its) leaves are very big.'

(Li and Thompson, 1989)

In the example above, ∅ represents the zero pronoun which exists because the topic, in this case 'the tree', was established at the beginning of the sentence and thus without the actual pronoun, one can understand the referent.

1.3 Pronouns in Japanese

Japanese pronouns, as compared to English and Mandarin Chinese, are a little different. They are restrictive in their uses and in many a times are omitted completely. Due to sociocultural factors, pronouns, especially personal pronouns are seldom used when referring to people. Instead proper nouns or names of the person are preferred. For example,

1d) *Watashi wa Nakagawa sensei ni piano o narai-mashi-ta. ****Kanojo*** *wa yuumeina pianisuto de, yoku shinbun ni mo not-te-i-mashi-ta.*
'I learned piano from Ms Nakagawa. **She** is a famous pianist, and often referred to in the newspaper.'

*Due to the status of Ms Nakagawa being the teacher of the speaker, '*kanojo* she' is incorrectly used and should be replaced by '*sensei* teacher' instead. The use of nouns instead of pronouns is preferred because using pronouns would make the speaker appear rude and disrespectful when referring to someone of respect such as a teacher. (Obana, 2000)

Unlike English pronouns, Japanese pronouns do not belong to a closed class of words (Backhouse, 1993). For example, the personal pronoun '*I*' in English, has several parallels such as '*watashi*' and '*watakushi*' (used by both males and females), '*ore*' and '*boku*' (used mainly by males) and '*atashi*' (used mainly by females). These are just a small fraction of words that can be used to represent the first person in the language. Depending on the dialect, the formality, politeness and most importantly, the social relationship between the speakers (Ono and Thompson, 2003), the personal pronoun '*I*' can be expressed in many different forms in Japanese. It goes the same for second and third person personal pronouns.

Furthermore, the so-called "pronouns" in Japanese did not started out as pronouns (Obana, 2000). Many of them came about from nouns in old Japanese, which had undergone semantic and pragmatic changes to become pronouns used in modern Japanese (Ishiyama, 2008). Take for example one of the forms of 'I', '*boku*' ('I' used mainly by males). It was previously used to mean 'servant' but has now evolved to a first person pronoun used in informal occasions such as when speaking to persons of the same or lower status (Ishiyama, 2008).

In addition, for the third person pronoun, '*kare*' and '*kanojo*' meaning 'he' and 'she', can also be used as a noun to refer to one's boyfriend or girlfriend as in lovers. This shows that pronouns in Japanese are not fixed as a grammatical class, unlike those in English.

For possessive pronouns in Japanese, like in Chinese, they are expressed through the addition of a possessive morpheme, the particle '*no*', to the personal pronouns, such as:

1e) *watashi* ***no***
'Mine'
1f) *kimi* ***no***
'Yours'
1g) *kare* ***no***
'His'
1h) *kanojo* ***no***
'Hers'

While for reflexive pronouns, in Japanese there is one main form - '*jibun*' which means 'self' and one can only interpret its referent through context as it does not differentiate by number.

1.4 The current study

As can be seen by the brief introduction of pronouns in the grammars of the three languages, pronouns being a form of language universals have "their inter-subjective and dialogic character hold a primacy over individual consciousness" (Violla, 2011). However, though they are, as their definition suggests, used across languages for similar purposes, they exist differently in terms of the number of types and how they are used or preferred.

According to Kim (2009), there exist qualitative and quantitative differences in the usage of the second person and first person plural pronouns in texts he examined from English and Korean newspapers. Texts pulled from an academic multilingual corpus such as the NTU Multilingual Corpus (NTU-MC) (Tan and Bond, 2011) seem to give similar results. Other than personal pronouns, other categories of pronouns seem to exist and used differently in different languages.

By analyzing pronouns of the three languages (English, Mandarin Chinese and Japanese) from three totally different language families (Indo-European, Sino-Tibetan and Japonic), this research intends to discover and describe the similarities and differences between their usage of pronouns and reasons for the differences. Using a corpus to do this allows for one to examine the distribution of pronouns in the source language, to contrast their use with their translated counterparts (Coussé and Auwera, 2012) and also to shed light on the characteristics of the individual languages (Wong, 2010). We expect the translated text (from English) in Mandarin Chinese and Japanese to have slightly more pronouns than their native text.

Despite this, there is surprisingly little research on comparisons of pronouns across languages and for those that study pronouns, most of them focus on the comparison of a specific type of pronoun such as personal pronouns on either a single language or on two languages. Furthermore, corpus-based studies on pronouns are seldom parallel, though it is increasingly becoming so (we will discuss past research in more detail in the next chapter). Hence, this study hopes to develop a greater understanding in the way pronouns are used across languages and also to contribute to the corpus research on crosslingual pronoun usage.

The current study thus sets to find out the qualitative and quantitative differences that exist in the pronouns of these three languages using a corpus. After introducing how pronouns work in the three languages in the first part of this paper, we go on to review past studies on the crosslingual comparisons of pronouns using non-corpus based research and also corpus-based studies in Chapter 2. Following after in Chapter 3, a description of the corpus used in the study can be found with details on how it was used to analyze the results, Next, Chapter 4 presents the results and in Chapter 5, we discuss the results. In the final chapter, Chapter 6, we arrive at a conclusion with suggestions for potential research.

CHAPTER TWO

2. LITERATURE REVIEW

This chapter will discuss about past studies that deal with pronouns across languages and also corpus-based research on pronouns.

Pronouns, being common across all languages, have interested many linguists and scholars to find out more about them across languages. Chen and Wu (2011) proved that although English plural personal pronouns are as Borthen (2010) claimed to be "referentially less well-behaved", this phenomenon is not limited to only plural personal pronouns as the Mandarin Chinese singular personal pronoun '他' *ta1* (it/he/she*) can also be less well-behaved and restrictive referentially. In their paper, by discussing the usage of singular 'they' in English, they supported Borthen's (2010) claim that plural personal pronouns are less restricted in their senses as they can also refer to a singular referent depending on the context and its usage is a pragmatic choice by the user (Chen and Wu, 2011). However, using examples of the Mandarin Chinese singular personal pronoun '他' *ta1* (it/he/she*), they showed that depending on the environment the pronoun is in, pragmatic interpretations of the pronoun allow it to represent plural entities as well (Chen and Wu, 2011). This argument opposed Borthen's (2010) study about plural and singular personal pronouns having dissimilar functions. From the two papers, we see that although pronouns seem easy to recognize, the environment they are in and our interpretation of their referents can differ greatly depending on context and pragmatic issues.

Kashima and Kashima (1998) did a large scaled project where 71 countries (71 cultures) and 39 languages were analyzed for their relationship between culture and language. The authors calculated the cultural scores, using Hofstede's, the Chinese Culture Connection's (CCC), Schwartz's and Smith and colleagues' (cited in Kashima and Kashima, 1998) cultural variables, 19 in total, inclusive of individualism, moral discipline, hierarchy, paternalism respectively and more. A literature survey was conducted to examine the main languages of the different countries and for the few countries with no literature available or where inconsistency arose, native speakers were interviewed (Kashima and Kashima, 1998). The cultural scores were then tabulated along with the languages' relation to personal pronoun use, particularly on first and second singular pronouns 'I' and 'you' and also the phenomenon pronoun drop. The number, as well as whether the pronouns can be dropped when they are the subject of sentences, were examined and they observed that cultures with pronoun drop in their language are generally more collectivistic than those without (Kashima and Kashima, 1998). Also, they found out that the semantics of the pronouns in question may determine the

different relationships between the speaker and the hearer (Kashima and Kashima, 1998). For second person singular pronouns, the number largely correlates with the social structure where languages with multiple second person singular pronouns tend to differentiate "self-other relationships" by the participants of the discourse (Kashima and Kashima, 1998). On the other hand, for first person singular pronouns, languages with only one first person singular pronoun have cultures that place more responsibility on the individual while those that have multiple first person singular pronouns do not (Kashima and Kashima, 1998). This study gave a huge insight on the fact that cultural differences can account for the differences in pronouns' existence and usage crosslingually.

A majority of corpus studies crosslinguistically focuses on the semantic areas of languages and there are few that concentrate on the grammatical areas, especially that of pronouns unlike monolingual corpus studies where both semantics and grammar are widely researched. For instance, Laitinen's (2007) book – '*Agreement patterns in English: Diachronic corpus studies on common-number pronouns*' (cited in Mair, 2009), shows the different monolingual corpora used for her studies on common-number pronouns. The British National Corpus (BNC) for example, helped her to find out about the usage of third person pronouns 'he' and 'they' in indefinite anaphora in written forms of English of today (Mair, 2009). From the BNC, Laitinen showed that the third person pronoun 'they' is increasingly being used for its neutral gender sense instead of 'he' in present day English where feminism is progressively getting stronger (Mair, 2009). Whereas the findings of pronouns and number in the Corpus of Early English Correspondence (CEEC) helped in discovering the existence of a "typological-diachronic drift" which cause English to lose its grammatical gender and almost its number agreement (Mair, 2009). Through the two different corpora, Laitinen realized the historical factors that affect the change in usage of pronouns in English.

Coussé and Auwera (2012) studied the human impersonal pronoun 'man' in Swedish and 'men' in Dutch using a Dutch-Swedish parallel corpus. This corpus contains the target languages in Swedish and Dutch with their translations respectively into the other language aligned at sentence level. For the Swedish texts, seven novels and four non-fiction texts were examined while five novels and one non-fiction text were investigated for Dutch. Findings showed that 'man' and 'men' have overlapping meanings but are used differently across the languages (Coussé and Auwera, 2012). Referents of 'man' can be either the speaker or another known person, whereas 'men' can only be used to refer to indefinite referents. In addition, while 'man' appeared often in the novels and non-fiction texts, 'men' appeared only in the non-fiction texts, showing the more restricted use of 'men' (Coussé and Auwera, 2012). Also, they found that for 'man' in Swedish, Dutch and English have similar personal pronouns as its correspondence, which is interesting because the impersonal pronouns for the

two languages are etymologically different. Other than pronoun correspondents, Coussé and Auwera (2012) also discovered a large number of substitute approaches for impersonal reference such as the agentless passive and non-finite clauses with 'to' infinitives. Despite their findings, the study was not reinforced by any grammatical resource such as ParGram (Butt et al, 2002) or The GF Resource Grammar Library (Ranta, 2009).

For differences between first person plural and second person pronouns in Korean and English, Kim (2009) derived qualitative and quantitative differences from his corpus study of texts from English and Korean newspaper science popularizations. Kim (2009) collected these texts from the online versions of two British newspapers – The Daily Telegraph and The Guardian and two Korean newspapers – Chosunilbo and Dongailbo. The texts are also taken from around the same period of time so that possible problems arising from changes of the genre in different periods could be avoided. Due to the difference in frequency of the published texts between the British and Korean newspaper and also the shorter lengths in Korean text as compared to British text, Kim (2009) had a greater number of Korean texts over a longer time period. He focused on "reader-involvement evoking (RIE)" acts of the article where the pronouns are claimed to be significant. After the comparison of the texts in both languages, Kim (2009) found out that both the first person plural and second person pronouns are lower in occurrence in the Korean texts and that in English, the first person plural and second person pronouns occurred at a similar frequency while the first person plural pronoun was employed more dominantly in Korean. The reasons for the difference in the distribution and choice of pronouns used in the two languages he believed was due to firstly, the linguistic features of Korean where agent omission is acceptable and secondly the sociocultural factors which have huge influences on the languages. In English, depending on the writers' stances, the different pronouns used can allow the writer to show solidarity with the readers or detaching himself from them. Conversely, in Korean, the more dominant usage of the first person plural instead of the second person pronoun was most likely due to its collectivistic society where indirectness is much preferred (Kim, 2009). However, Kim's (2009) study is not parallel and is not backed by a grammatical resource as well.

The same applies to Smith's (2004) study on personal pronouns and possessive determiners in advertising texts using a English-Russian parallel corpus. Smith (2004) used two corpora for her research where one of them is the English-Russian corpus with forty-five advertisement texts in English as the source language and their Russian translations and a monolingual Russian corpus with twenty-three advertisement texts in Russian acting as a control. Through analyzing the corpora, she assessed her results in three categories – consumers, advertiser's company and intermediary. She realized that the different personal pronouns, used in English to establish the relationships between consumers, the advertiser's

company and an intermediary, operate similarly in Russian for the Russian audience in terms of consumers but not for the companies and intermediaries (Smith, 2004). Since the second person pronoun 'you' is significant in building relationships with the consumers, it is used frequently in the English texts and likewise in its Russian translation, although there exist a formal and familiar variant for the same pronoun, both forms are used frequently (Smith, 2004). Also, for possessive determiners, both the Russian translations and the monolingual Russian advertisement texts used them similarly to the English texts although in Russian, possessives are unnecessary if ownership is established (Smith, 2004). However, in terms of the representation of the company, the first person plural pronoun 'we' was used more often in the English texts than in its Russian translations. Furthermore, the monolingual Russian corpus displayed the use of inclusive 'we', which English does not have. This was likely due to the collectivism culture in Russia (Smith, 2004). Lastly, with regards to the intermediaries, the first person pronoun 'I' was rarely found in the Russian translations and never in the monolingual corpus. This difference is also said to result due to the collectivism culture in Russia where expression of individualism is avoided (Smith, 2004). Smith's (2004) study shows that the use of pronouns can indeed differ in quantity and quality due to the different cultural backgrounds of the language in a genre different from Kim's (2009) study above. However, there was no back up from any grammatical resource. But, Smith (2004) compared her results with another corpus, which is monolingual in Russian to confirm their validity.

Wong (2010), motivated by Baker's 1992 study (cited in Wong, 2010) which found that there is almost no grammatical category that is consistently uniform across different languages, used the Babel English-Chinese Parallel Corpus to investigate whether the above is true by examining existential constructions in English-Chinese translations. The Babel English-Chinese Parallel Corpus consists of 327 English texts and their Chinese counterparts, totaling 550,000 tokens with 250,000 in English and 300,000 in Chinese. Wong (2010) focused on existential *there*-constructions for her study and extracted 368 examples from the corpus. From her study, she found that the existential constructions used in English, for a greater extent, do not exist in Chinese and in places with complex notional subjects, Chinese tend to reorder and restructure the constructions, thus resulting in no direct correspondences of Chinese for the English source texts (Wong, 2010). To validate her results, Wong (2010) substantiated her findings with traditional Chinese grammars and through her study, she found that Baker's statement holds true.

For the few corpus studies on pronouns crosslinguistically, they are usually not parallel and/or not supported by any concrete grammar resource or database such as Wordnet (Bond et al, 2013), ParGram (Butt et al, 2002) or The GF Resource Grammar Library (Ranta, 2009). Otherwise, they mainly deal with European languages or other grammar features (as seen

from the studies above). Another factor to take note is that these studies usually deal with a small, fixed set of pronouns such as specifically personal pronouns only or sometimes even just a subset of that, such as those that focus on the first and second personal pronouns.

Therefore, the current study hopes to fill up these research gaps by first studying other categories of pronouns other than personal pronouns. And secondly, by analyzing the pronouns with a parallel corpus, crosslinguistically, specifically Asian languages in contrast with an Indo-European language (English). In particular we will look at how pronouns are translated (or not translated).

CHAPTER THREE

3. METHODOLOGY

This section introduces the corpus and describes how the corpus data was annotated and analyzed on the monolingual and crosslingual levels.

3.1 The Corpus

This project uses corpora as the base of the study. The corpus used is the NTU Multilingual Corpus (NTU-MC) that is currently still being developed (Tan and Bond, 2011). The NTU-MC exploits the linguistic diversity available in Singapore for the collection of a vast variety of texts from different languages (Tan and Bond, 2011). The current version is an annotated collection of around 26,000 sentences (~595,000 words) in 7 languages (Arabic, English, Mandarin Chinese, Japanese, Korean, Indonesian and Vietnamese) from 7 language families (Afro-Asiatic, Indo-European, Sino-Tibetan, Japonic, Korean (language isolate), Austronesian and Austro-Asiatic) (Tan and Bond, 2011). Two kinds of annotation are applied in the NTU-MC – monolingual annotation where texts are tagged for parts of speech (POS) and sense and crosslingual annotation where texts are aligned across sentences (Bond et al, 2013, Bond & Wang 2014).

3.2 Corpus Data

Pronouns from the three languages (English, Mandarin Chinese and Japanese) were extracted from four data sets in the NTU-MC. They are two short stories from Sherlock Holmes – *The Adventure of the Speckled Band* and *The Adventure of the Dancing Men*, an essay named *The Cathedral and the Bazaar* and on-line articles about Singapore tourism. In each set, English is the source language while Mandarin Chinese and Japanese translation texts are aligned to it at the concept level. The texts have been tokenized and automatically POS tagged.

3.3 Componential analysis of pronouns

The first stage was to extract pronouns using the POS mapping and analyze them componentially. The pronouns were separated into nine categories namely - Head, Number, Gender, Case, Type, Formality, Politeness, Distance from Speaker and lastly Honorific. The features chosen are in line with other research and reference grammars (Backhouse, 1993; Carter & McCarthy, 2006; Collins Cobuild English grammar, 2005; Li & Thompson, 1989; Obana, 2000; Ross & Ma, 2006; Sun, 2006; Yip & Rimmington, 1997), which have previously examined pronouns. The purpose of this componential analysis is to code the pronouns so that we can compare and contrast them across languages. This also allows the auto-tagging programme to recognize and link the pronouns by their code. This stage took around two weeks due to the detailed componential analysis of every pronoun in the four

subcorpora and analyzing ambiguous forms particularly in Japanese. Below is a table showing the different features under each heading (Table 3.1).

Head	**Number**	**Gender**	**Case**	**Type**	**Formality**	**Politeness**	**Distance from Speaker**	**Honorific**
D – Demonstratives	**D –** Dual	**F –** Feminine	**O –** Objective	**A** - Assertive Existential	**F** - Formal	**P** - Polite	**D** - Distal	**A** - Archaic Honorific
E – Entity	**P –** Plural	**M –** Masculine	**P –** Possessive	**E** - Elective Existential	**I** - Informal		**M** - Medial	**H** - Honorific
T – Time	**S** - Singular	**N** - Neuter	**S** - Subjective	**N** - Negative			**P** - Proximal	**X** - Non-honorific
M – Manner				**O** - Other				
P – Person				**R** - Reciprocal				
L – Place (Location)				**U** - Universal				
Q - Quantifier				**I** - Interrogative				
O – Thing (object)				**S** - Reflexive				

Table 3.1: The 9 types of pronoun features

In Table 3.1 shown above, we can see the nine different categories that were used to analyze the pronouns. In the first column - Head, there are altogether eight components. They are Demonstratives, Entity, Time, Manner, Person, Location, Quantifier and Object. Every pronoun extracted will be tagged with one of these features. For example, Demonstratives refer to pronouns such as 'this' and 'that' while Entity are pronouns that do not have a specific category of referent, as it can refer to both person and object. Such pronouns are 'all', '俩 *liang3* both' and 'いくつ *ikutsu* some'. 'When', 'how' and 'where' are examples of pronouns labeled under Time, Manner and Location respectively. 'Little' and 'many' are labeled as Quantifier pronouns. For English pronouns, words that end with '~thing' are grouped under Object, while for Mandarin Chinese and Japanese pronouns, they are not so clear-cut. Lastly personal pronouns and pronouns that talk about people like 'everybody' and '自己 *zi4ji3* self' are categorized under Person.

The next column - Number is where we differentiate the pronouns that mark for number. For this feature, we identified three kinds of number – Dual, Plural and Singular. 'Both' is an example of Dual, 'those' for Plural and '这 *zhe4* this' for Singular.

For the third column - Gender, three features were identified as well – Masculine, Feminine and Neuter. 'It' in English is a neuter pronoun while '她 *ta1* she' in Mandarin Chinese is Feminine and 'ぼく *boku* I' in Japanese is Masculine.

Following after, we will check if the pronouns are marked for Case (fourth column). In Case, there are Subjective, Objective and Possessive pronouns. Only English separates for all three cases, while Mandarin Chinese has third person, singular pronouns '之 *zhi1*' and '其 *qi2*' for possessive and Japanese has none.

The next column – Type, differentiates the pronouns by Assertive Existential, Elective Existential, Negative, Reflexive, Reciprocal, Universal, Interrogative or Other. Pronouns with 'some ~' are assertive existential as they are unspecified measures of entity, while 'any~' are elective existential because they can be either some or none, Negative are pronouns such as 'none', while pronouns with the meaning of 'self' are reflexive. Reciprocal are pronouns that mark for each other such as '彼此 *bi2ci3*' in Mandarin Chinese. Universal consists of

pronouns that indicate all and every. Interrogative are question words and the remaining pronouns belong to Other.

The sixth column shows Formality, whether the pronouns are informal or formal. This is mainly for the Japanese pronouns, which mark for formality. The same goes for Politeness where it comprises pronouns that show politeness such as '您 *nin3* you' in Mandarin Chinese.

The eighth column - Distance from speaker, is meant for pronouns that mark for Proximal, Medial or Distal distance from the speaker. These pronouns are used for Location pronouns such as 'そこ *soko* there' which tells both the distance from the speaker and distance from the recipient.

Lastly, the Honorific component is for honorific pronouns. Only Japanese pronouns are differentiated here. Tables 3.2-3.4 below show some examples of how the pronouns are being categorized into their respective components.

Type / Head	Assertive Existential	Elective Existential	Negative	Other	Reciprocal	Universal	Interrogative	Reflexive
Entity	有的 / some / いくつ	Any / either	None / neither	Another / その他		各　自 / everything / あらゆる	什么样 / どれ	
Time							何时 / いつ	
Manner							怎样	
Person	Somebody	Anyone / anybody		对方	彼此	大　家 / everybody / みなさん	いずれ / だれ	本身 / 自己 / himself / myself / themselves / yourself / おのれ
Place (Location)				よそ		处处	何处 / 哪里 / どこ / どちら	
Thing (Object)	Something	Anything				一　切 / everything	什么 / なに / どちら	

Table 3.2: Examples of pronouns that are categorised

Demonstratives

Distance from speaker / Number	Proximal	Medial	Distal
Singular	这 / 此 / this / この / これ	那 / that / その / それ	あの / あれ
Plural	这些 / these / これら	those / それら	

Table 3.3: Categorising demonstratives

Quantifiers

Number	**Singular**	**Plural**
Pronouns	little	many

Table 3.4: Categorizing quantifiers

3.4 Auto-tagging and Manual tagging of pronouns

After analyzing the pronouns by their different components, the second stage was using an auto-tagging programme to run through them and tag them according to the componential analysis to the texts in the corpora at the monolingual annotation level. This is needed because the existing Wordnets do not have synsets for pronouns. After which, the pronouns are then linked together via crosslingual annotation by aligning the pronouns from the source language, English, to Japanese and to Mandarin Chinese at the sentence level. However, due to time and space constraints, for this study, the manual tagging will only focus on one subcorpus, 'The Adventure of the Speckled Band' of the Sherlock Holmes short stories. At the crosslingual annotation level, the pronouns were checked manually to see if they are tagged as pronouns correctly by the auto-tagging programme and whether the concept links between the source language and target language are accurate. This was done several times to ensure accuracy. This stage took around four weeks to complete both English-Chinese and English-Japanese corpora, with a longer time needed for the English-Chinese one due to the greater number of pronouns present there. On average, three to four sentences can be done every hour.

The programme links the pronouns of the two languages together if at least five of the pronoun features (refer to Table 3.1) are matched. For example,

'*she*' has a tag 77000044-n which equates to a 3rd person, singular, feminine and subjective pronoun and a code <3:S:F::S::::> (refer to table above) can be linked to '*她* ta1' which has a tag 77000041-n equating to a 3rd person, singular, feminine pronoun and a code <3:S:F::::::> in the English-Chinese corpus and also to '*彼女* kanojo' which has a tag 77000042-n, meaning a 3rd person, singular, feminine, formal pronoun with the code <3:S:F:::F:::> in the English-Japanese corpus.

The two pronouns in Mandarin Chinese and Japanese above are matched once the first five codes are the same and then linked up respectively to the source language through the auto-tagging programme.

Although there are many pragmatic and social differences among the three languages, such as Japanese being the only language, which consists of different speech levels, and can thus prevent the pronouns from linking, we decided that by including all of the headings, we might be able to see clearer how the pronouns could be linked together at the sentence level. Therefore, we have decided to over−link the pronouns by matching at least five of the nine features so that we can get a more detailed result of the linkages.

Through the auto-tagging programme, the pronouns that are matched for at least five codes, are linked with '~' relation which are for pronouns that have related meaning but are not identical. After working through the texts, the pronouns that are deemed identical to each other at the sentence level have their concept link changed to '=' relation which meant that they are used for the same purpose and function in their

respective language. In the example below, '*I*' is lexically and translation wise equivalent to '我 *wo3*' and thus are linked with '='.

3a) English: **I** shall look forward to seeing you again this afternoon
Chinese: **我** 盼望 下午 能 再见 到 你们。
Wo3 pan4wang4 xia4wu3 neng2 zai4jian4 dao4 ni3men2
'I hope to be able to see you again in the afternoon.'

Another concept link that was used is ':' relation. Pronouns that are linked with ':' are pronouns that have the same meaning across translation but is itself lexically different from the other. For example as shown below in (3b),

3b) English: **It** is a swamp adder!
Chinese: **这** 是 一 条 沼地 蝰蛇!
Zhe4 shi4 yi1 tiao2 zhao3di4 kui2she2
'This is a swamp adder!'

'It' and '这 *zhe4* this' are used to refer to the same thing in its respective language but lexically it is not equivalent – '*it*' has a tag 77000053-n meaning it is a 3rd person, singular, neuter pronoun with the code <3:S:N::::::> while '这 *zhe4* this' has a tag 77000061-n which equates to it being a demonstrative, singular and proximal pronoun with the code <D:S::::::P:>. Since the first five codes do not match, the two pronouns were previously not linked after the run through of the auto-tagging programme. However, by inferring from the contextual meaning, the two pronouns are actually translation equivalents.

The tagging guidelines used above are taken from Mok's (cited in Gao, 2012) and Gao's (2012) paper where it was used as a base for the establishment of the relationships between the tokens in the source language and its counterparts (Bond et al. 2013, Bond & Wang 2014). A table showing the brief summary of the symbols used to tag the relationships is shown below:

Symbol	Relationship
=	Identical meaning (Translation wise and lexical wise)
~	Similar or related meaning (Translation wise and lexical wise)
:	Translation equivalent but not lexically the same

Table 3.5: Summary of symbols used for linking the pronouns

3.5 Revision of annotations and tagging

During the manual checking of the pronouns across the sentence pairs in both the English-Chinese and English-Japanese translated texts, errors in the tagging of the pronouns and also their concept linking from source language to its translated counterparts were found. For the last stage, we corrected these errors as we went through the texts. Another two weeks were spent for this stage, using double the speed for the previous stage with six to eight sentences being revised per hour.

On the monolingual level, there were two kinds of error –
1. Words that are supposed to be pronouns were not tagged
    a. *there*: Julia went ***there*** at Christmas two years ago

b. *你们*: 我　一直　在　焦急　地　盼　着　***<u>你们</u>***
wo3 yi4zhi2 zai4 jiao1ji2 de4 pan4 zhe4 ni3men
'I have been waiting for you anxiously.'

c. ***我が***: ***<u>我が</u>*** 友　シャーロック·ホームズの手法　をここ
Waga tomo sharokku homuzu no shuho o koko
八年記　録　している手帳 に 記さ れた
hachinen kiroku shite iru techo ni shirusa reta
'I have recorded my friend Sherlock Holmes' eight years of methods in this notebook here.'

In the above three examples, the words that are underlined, function as pronouns in their respective sentences, however, they were not tagged as pronouns, showing that there is an error in the auto-tagging as it missed out on obvious pronouns.

2. Words that are not supposed to be pronouns are tagged

a. Dummy *it:* ***<u>It</u>*** was a perfect day, with a bright sun and a few fleecy clouds in the heavens.
b. Pronouns in idiomatic phrases: "***<u>My</u>*** God!" I whispered, "Did you see it?"

These underlined words are non-referential pronouns. In example (a), 'it' is a dummy pronoun and thus does not refer to any particular noun and in (b), pronouns used in idiomatic phrases do not actually have any particular referent as well since they are always used in the same way in these phrases. However, the auto-tagging programme does not distinguish referential from non-referential pronouns (except for existential there) through semantic differentiation yet and thus through just POS tagging, these words were not corrected as non-referential pronouns.

On the crosslingual level, there were also two kinds of error –
1. Pronouns with the same or similar meanings that are supposed to be linked are not,

a. *that* to *那*
English: "Yes, ***<u>that</u>*** is the Crown."
Chinese: "是的，***<u>那</u>*** 是　克朗　　旅店。"
Shi4de, na4 shi4 ke4lang3 lv3dian4
'Yes, that is Crown Hotel.'

b. *he* to *彼*
English: ***<u>he</u>*** refused to associate himself with any investigation which did not tend towards the unusual, and even the fantastic.
Japanese: ***<u>彼</u>*** は 異常な、さらに は 奇想天外な
Kare wa ijona, sarani wa kisotengaina
調査でなければ、関わる　の を 拒んだからである。
chosadenakereba, kakawaru no o kobandakaradearu.
'This is because he refused to be involved in further investigation that are not unusual or out of this world.'

The pronouns in the two examples above have the same meaning and the same referent with their translated counterparts. Yet, they were not linked with each other as having the same conceptual link.

2. Wrong conceptual links of pronouns in the source language to other similar pronouns in the translated texts that exist in the same sentence

English: and then withdraw quietly with everything which ***you*** are likely to want into the room which you used to occupy.

Chinese: 随 后 带 上 你 可能 需要 的 东西， 悄
Sui2 hou4 dai4 shang1 ni2 ke3neng2 xu1yao4 de dong1xi, qiao1
悄 地 回到 ***你*** 过去 住 的 房间。
qiao1 de hui2dao4 ni3 guo4qu4 zhu4 de fang2jian1

The two pronouns that were underlined were linked to each other after being run through by the auto-tagging programme. Although as can be seen from the sentences, the referent for the pronoun is the same one, the environment in which it is in shows that it is not a suitable match for the linked pronoun in the translated text. The first ‘you’ should be linked to the first ‘你 *ni3*’ and the second one to the underlined ‘你 *ni3*’.

CHAPTER FOUR

4. RESULTS

Having linked and tagged the relationships between words, we proceeded to count the number of pronouns in each language and their links. The results shown are solely based on one subcorpus – *The Adventure of the Speckled Band*. Below is the table showing the number of pronouns found in each language that we labeled as 'true' and 'false' pronouns where 'true' are pronouns that are referential while 'false' pronouns refer to the pronouns that are non-referential and are not the pronouns we are looking for.

Language	English	Mandarin Chinese	Japanese
No. of False Pronoun	75	19	51
No. of True Pronoun	1370	1177	463
Total no. of pronoun	1445	1196	514

Table 4.1: Number of pronouns found in the two corpora

Out of a total of 3155 pronouns found in the corpus, 94.8% is 'true' in English, 98.4% is 'true' in Mandarin Chinese and 90.1% is 'true' in Japanese. As shown, among the three languages, the English text has the most number of pronouns, followed by the Mandarin Chinese text and then lastly the Japanese text but Mandarin Chinese has the highest percentage of referential pronouns, followed by English and then Japanese.

The results for the linkage of the pronouns are separated into two parts for better understanding – the first part being the results for the English-Chinese corpus and the second part for the results found from the English-Japanese corpus. Below is a summary for the number of links found for the pronouns in the English-Chinese corpus.

	Linked Pronouns						Non-Linked Pronouns	
	No. of features matched					Pronoun – Non-Pronoun	English	Mandarin Chinese
	5	6	7	8	9			
No. of pronoun	5	19	54	789	58	134	369	215

Table 4.2: Summary of the linkage of pronouns in the English-Chinese corpus

There are in total 925 English to Chinese pronouns linked to each other, with 0.5% of them having only 5 pronoun features match, 2.1% having 6 pronoun features match, 5.8% having 7 pronoun features match, 85.3% having 8 features match and 6.3% having 9 pronoun features match where 9 is the maximum match. The respective pronoun features can be seen in Table 3.1 in the Methodology section. The majority of the linked English-Chinese pronouns have 8 out of 9 matched features.

There are also 134 pronouns that are linked to non-pronouns. 76 of them are English pronouns while 58 of them are Chinese pronouns. Out of the 1370 'true' English pronouns, 26.9% of them are not linked. For the Chinese 'true' pronouns, only 18.2% were not linked to anything.

Also, according to the numbers, the bulk of 'true' pronouns are linked with 8 matched features, followed by pronouns which can only be found in one language whereby this happens more frequently in the English source text as compared to its Mandarin Chinese counterpart. After which, there was more than double for the pronouns that

were found to match non-pronouns compared to pronouns which have 9 features matching which preceded the pronouns having 7 matches, 6 matches and 5 matches respectively.

In the English-Japanese corpus, the table below provides a summary for the number of links found for the English and Japanese pronouns.

	Linked Pronouns						Non-Linked Pronouns	
	No. of features matched					Pronoun – Non-Pronoun	English	Japanese
	5	6	7	8	9			
No. of pronoun	15	120	114	37	32	139	943	109

Table 4.3: Summary of the linkage of pronouns in the English-Japanese corpus

There are in total 318 linked English to Japanese pronouns. Out of these, 4.7% have 5 matched features, 37.7% have 6 matched features, 35.8% have 7 matched features, 11.6% have 8 matched features and 10% have 9 matched features. The majority of the linked English-Japanese pronouns, unlike the English-Chinese corpus, have around 6 to 7 matched features.

Similar to the English-Chinese corpus, there are 139 pronouns in the English-Japanese corpus that are linked to non-pronouns with English being more common for this occurrence. 109 of the pronouns are English pronouns and the other 30 are Japanese pronouns. In contrast to the English-Chinese corpus, the English pronouns in the English-Japanese corpus that are not linked to anything accounts for 68.8% of the 'true' English pronouns. For the Japanese pronouns, 23.5% of them are not linked to any English words in the English source text.

From the numbers seen, most of the pronouns in the English text are not linked, meaning that they have no Japanese translation in the Japanese text. Comparable to the English-Chinese corpus, the number of pronouns linked to non-pronouns in the English-Japanese corpus has the second highest number. Pronouns, which have a match of 6 features, came next, followed by pronouns that match 7 out of 9 features. There are fewer Japanese pronouns that are not linked to any in the English source text than those that match around 6-7 features. Pronouns that match almost completely (8 out of 9 features) or those that match for all 9 features are infrequent. The least of the pronouns checked, have matches of 5 features with each other.

Comparing the results found in the English-Chinese corpus and those in the English-Japanese corpus, there are a few similarities and more differences. The only two similarities are that both have around the same number of pronouns that participated in linkage to non-pronouns and have the least number of pronouns that match for 5 features. Other than these two similarities, the results found in the English-Chinese corpus and the English-Japanese corpus are extremely different.

CHAPTER FIVE

5. DISCUSSION

In this section, we will account for the differences in the number of pronouns found in each language, discuss how they are linked in their respective corpus and also review some interesting finds.

5.1 English has the most pronouns, followed by Mandarin Chinese and lastly Japanese

As shown in the results above in Table 4.1, out of the three languages, English has the most number of pronouns, followed by Mandarin Chinese and then Japanese. However, Mandarin Chinese has the highest percentage of referential pronouns, followed by English and then Japanese. This is because English tend to use pronouns for more non-referential purposes as compared to Mandarin Chinese such as having dummy 'it', existential 'there' and also complementizers like 'that' and 'which'. Also, in English, many pronouns can also double up as determiners (Collins, 2005). Determiners share many common words with pronouns such as 'this', 'that' and indefinite ones such as 'all' and 'some'. Whereas for Japanese, as stated above in the introduction, some of the pronouns have other meanings and at times can be used as nouns. Also, being an agglutinative language, Japanese marks their verbs (Backhouse, 1993) and different particles can be used with a pronoun to produce different meanings altogether which often do not function as pronouns anymore.

5.2 Differences in the pronoun linkage in both corpora

Between the English-Chinese corpus and the English-Japanese corpus, there were more differences than similarities. They differ in the number of linked pronouns and also in the number of pronouns that are not linked. The first major difference found in the two corpora is the percentage of English pronouns that have no translation. In the English-Chinese corpus, majority of the English pronouns have translation in the Chinese text but in the English-Japanese corpus, majority of the English pronouns have no translation in the Japanese text. We expected the result for the English-Japanese corpus but not for the English-Chinese corpus. This difference in number can be caused by many reasons, some of which are discussed as follow.

Firstly, the way in which the three languages use pronouns can be a major factor in determining that result. English uses pronouns frequently (Collins, 2005). Pronouns in English can take the place of nouns in noun phrases and also to indicate the subject and object of the sentence (Carter and McCarthy, 2006). Since English is an isolating language, it seldom relies on inflection to give us the information we need to make sense of the referents the pronouns represent. Thus, a greater amount of pronouns is required in English to be able to properly express the referent. Other than personal pronouns having many different forms, English has other categories of pronouns that both Mandarin Chinese and Japanese do not have. For example, for the component Negative, English has 'none' and 'nothing' which do not have identical correspondents in Mandarin Chinese and Japanese. This is because both languages tend to use verbs to express negativity instead of marking it in the pronoun like in (5a).

5a) English: but **none** commonplace
Chinese: 但是　却　没有　一 例 是 平淡无奇　的
Dan4shi4 que4 mei2you3 yi1 li4 shi4 ping2dan4wu2qi2 de
'But, there is no one case that is boring.'
Japanese: どれ　も　尋常で　はない　事件である
Dore mo jinjode wanai jikendearu
'There is no unusual incident.'

In addition, Mandarin Chinese and Japanese are topic-prominent languages (Li and Thompson, 1989; Obana, 2000). Once the topic is established, sentences following it omit any pronouns, as there is no need for them to refer back as the readers can infer from contextual knowledge the subject of the sentence. The following example (5b) shows this:

5b) English: ***My companion*** sat in the front of the trap, ***his*** arms folded, ***his*** hat pulled down over ***his*** eyes, and ***his*** chin sunk upon ***his*** breast, buried in the deepest thought.
Chinese: ***我 的 伙伴*** 双　臂 交叉　地 坐　在 马车　的
Wo3 de huo3ban4 shuang1 bi4 jiao1cha1 de zuo4 zai4 ma3 che1 de
前部，∅ 帽子 耷　拉 下 来 遮住　了 ∅ 眼睛，
qian2bu4, ∅mao4zi3 song3 la1 xia4 lai2 zhe1zhu4 le4 ∅yan3 jing1,
∅ 头 垂　到 ∅ 胸　前，深深　地 陷 入 沉
∅tou2 chui2 dao4 ∅xiong1 qian2, shen1shen1 de xian4 ru4 chen2
思 之 中。
si1 zhi1 zhong1.
'My companion folded (his) arms and sat in the front of the horse-drawn carriage, (his) hat pulled down to cover (his) eyes, (his) head drooped to the front of (his) chest, deeply in thought.'
Japanese: ***我が友*** は、軽二輪 馬車 の 前 に 座って∅ 腕組み をし、
Waga tomo wa, keinirin basha no mae ni suwatte ∅udegumi o shi,
∅ 帽子を ∅ 目深　に かぶり、∅顎 を∅ 胸 に うずめて考え
∅boshi o ∅mabuka ni kaburi, ∅ago o ∅mune ni uzumete
込んでいた。
kangaekonde ita.
'My friend has (his) arms folded, sitting in front of a light two-wheeled horse-drawn carriage, wearing (his) hat over (his) eyes, chin buried in his chest, deep in thought.

As can be seen in (5b), although English has stated the topic or the subject at the start of the sentence, it requires the possessive pronoun 'his' to refer back so that the readers will know whose arm, whose hat, whose eyes etcetera. However, this is not necessary in Mandarin Chinese and Japanese. With the topic stated at the start of the sentence, the following noun phrases do not require a possessive pronoun to indicate its possessor.

Therefore, we did not expect the Mandarin Chinese text to have that many pronouns since it is a topic-prominent language, but maybe due to the influence from the English source text, the translator chose to employ more pronouns in the Mandarin Chinese translation where it would be just as grammatically correct without them. As seen in (5c) below, the number of pronouns in the English text and its Mandarin Chinese correspondents are equal. However, the two first person pronouns after the first clause in the Mandarin Chinese translation are actually not needed. Without these pronouns, the sentence would still be grammatically correct (^Chinese).

5c) English: As ***I*** opened my door ***I*** seemed to hear a low whistle, such as ***my*** sister described

Chinese: 就 在 ***我*** 开启 房 门 时，***我*** 仿佛 听 到 一
Jiu4 zai4 wo3 kai1qi3 fang2 men2 shi2, wo3 fang3fu2 ting1 dao4 yi4
声 轻轻 的 就 象 *我* 姐姐 说 的 那 样 的
sheng1 qing1qing1 de jiu4 xiang4 wo3 jie3jie shuo1 de na4 yang4 de
口哨 声
kou3shao4 sheng1
'Just when I was opening the room door, I seem to hear a soft whistle just like what my sister said.'

^Chinese: 就 在 ***我*** 开启 房 门 时，***Ø*** 仿佛 听 到 一
Jiu4 zai4 wo3 kai1qi3 fang2 men2 shi2, ***Ø*** *fang3fu2 ting1 dao4 yi4*
声 轻轻 的 就 象 *Ø* 姐姐 说 的 那 样 的
sheng1 qing1qing1 de jiu4 xiang4 ***Ø*** *jie3jie shuo1 de na4 yang4 de*
口哨 声
kou3shao4 sheng1
'Just when I was opening the room door, (I) seem to hear a soft whistle just like what (my) sister said.'

Despite this, Mandarin Chinese and English do have similar pronouns that do not exist in Japanese. For example in (5d), the Entity, Universal pronoun 'all' has '一切 *yi1qie4*' as its counterpart in the Mandarin Chinese text, but not in the Japanese text.

5d) English: and then ***all*** was silent once more
Chinese: 接着， ***一切*** 又 都 沉寂 下来
Jie1zhe, yi1qie4 you4 dou1 chen2ji4 xia4lai2
'Then, all was silent again.'
Japanese: また 静か に なり
Mata shizuka ni nari
'(It) became quiet again.'

Furthermore, out of the three languages, only Japanese marks politeness and some evidentiality on the verb (Backhouse, 1993), making the use of pronouns rather unnecessary and this seems to play an important role in determining the much fewer pronouns found in the corpus as compared to the English source text and Chinese translation text, resulting in the low rate of links to the English pronouns in the original text. One example can be seen below in (5e):

5e) English: ***I*** have heard of ***you***, Mr. Holmes

Japanese: ***あなた*** の こと は、以前 から お 聞き して います。

Anata no koto wa, izen kara o kiki shite imasu.

'About you, (I) humbly heard previously.'

In the English text in (5e), there were two pronouns in the sentence. However, in the Japanese translation, only one pronoun was found which is '**あなた** *anata*' in correspondence to 'you' in the English text. The 'I' in Japanese is not needed as it is encoded in the verb '聞きして います *kiki shite imasu*'. This verb is a polite and formal way of saying 'I have heard' although it literally translates to 'humbly heard'. In this case, there is no need for the first person pronoun as the verb is used in a situation where the speaker respects the recipient and thus if 'I' was used, it would seem too rude and not respectful as one of the politeness strategy in Japanese is to "avoid one's direct reference to oneself" (Obana, 2000). Also, Japanese is a "speaker-oriented" language (Obana, 2000), which means that constructions are done from the speaker's perspective and it is not necessary to refer to 'I'. With the incorporation of the meaning of 'I' in the verb, there is hence no need to use the pronoun in Japanese to indicate the subject and at the same time, the speaker appears polite and respectful.

Between the English-Chinese corpus and the English-Japanese corpus, another major difference is the number of corresponding features that majority of the linked pronouns have. For the English-Chinese corpus, majority of the linked pronouns have 8 matching pronoun features while for the English-Japanese corpus, majority of the linked pronouns have around 6 to 7 matching pronoun features. This is most likely due to Japanese language having different speech levels (Obana, 2000). The different speech levels cause a differentiation between the pronouns, resulting in Japanese having a few different words for the same pronoun. For example, for the first person pronoun, in Japanese there are variations such as 'わし *washi*' which also marks for masculine and informal and '私 *watashi*' which marks for formal and politeness. These features do not exist in English but from the perspective of semantics, they should be linked to the first person pronouns in English. This problem does not exist in Mandarin Chinese, as there is no such differentiation in speech levels in Mandarin Chinese. Therefore, more features can be matched.

Also, from the linking of the pronouns, there were many cases where English pronouns were linked to Mandarin Chinese and Japanese pronouns that are different in meaning such as the third person pronoun 'it' in the English text to the demonstrative pronoun 'そこ *soko* that' in Japanese. Although this happens in the English-Chinese corpus as well, they are less frequent, thus resulting in more of the pronouns linked have more matched features as compared to those in the English-Japanese corpus.

5.3 Deprominalisation occurs almost evenly in both the English-Chinese and English-Japanese corpora

As seen in the results, the number of pronouns matched to non-pronouns in the English-Chinese corpus is around the same. This result is not expected as deprominalisation was predicted to occur much more frequently in the English-

Japanese corpus than in the English-Chinese corpus. This is because according to Maynard, 1990 (cited in Gao, 2012) the use of pronouns is much more uncommon in Japanese. Also, honorific forms, respective and polite ways of using the language is highly valued in the culture. Therefore, to avoid being rude and disrespectful, for example, instead of using pronouns to refer to people, usually their names or their titles are preferred. An example (5f) can be seen below, where 'he' was translated into 'Holmes':

5f) English: ***He*** had ceased to strike and was gazing up at the ventilator

Japanese: ***ホームズ***は 打ちつける の を 止めて、通気口 を 見上げた

Homuzu wa uchitsukeru no o tomete, tsukiguchi o miageta

'Holmes stopped striking and looked up at the vents'

In (5f) above, the pronoun 'he' is linked to the name 'Holmes' instead of a pronoun. However, since deprominalisation is rather equal for both corpora, it may be because the Japanese translation was influenced by the English source text and thus used more corresponding pronouns than usual to match the English source text.

5.4 Interesting cases found

From the tagging of the pronouns and their concept links, there were a few interesting cases that were found. In the English source text, we realized that pronouns often exist in idiomatic phrases. However, these pronouns do not actually have any particular antecedent to refer to as they are almost always used in the same way regardless of its environment and this causes the linking of the pronouns to be somewhat problematic.

5g) English: ***My*** God!

Chinese: 天 哪！

Tian1na

'Heaven!'

Japanese: なんてこったい！

Nantekotta i

'What the heck'

The above example shows that 'my' is used here as a pronoun in an idiomatic phrase and after translation, no pronouns were seen. In both the Mandarin Chinese and Japanese text, their own version of an exclamation phrase here proves that pronouns in idiomatic phrases cannot be easily defined.

Also, when translated to Mandarin Chinese or Japanese, usually there would be no exact corresponding pronouns as the phrases are translated according to their idiomatic meaning and not the literal meanings of the words.

5h) English: It is very kind of you.

Chinese: 非 常 感 谢！

Fei1chang2gan3xie4

'Very grateful'

Japanese: 感謝　して いるよ
Kansha shite iru yo
'I'm grateful.'

As seen in the example above, both the Mandarin Chinese and Japanese translated the English source text in different ways. For the Mandarin Chinese text, its literal meaning is 'very grateful' and can be understood as 'thank you very much' as well. Whereas for the Japanese translation, it means 'I'm grateful'. Although 'grateful' was not inside the English text, the translators of the other two target languages took the figurative meaning and translated according to that. Another example is shown below:

5i) English: I assure you that ***I*** am in ***your*** hands.
Chinese: 我　向　你　保证，　***我***　一切　听从　***你***　的　吩咐
Wo3 xiang4 ni3 bao3zheng4, wo3 yi2qie4 ting1cong2 ni3 de fen1fu4.
'I promise you, I will obey all your instructions'
Japanese: ***あなた*** の 手 に すべて を お ゆだね します わ
Anata no te ni subete o o yudane shimasu wa
'I will leave everything in your hands'

The Chinese translation here again chose to take the figurative meaning from the English text (of the underlined phrase) and translated it to 'I will obey all your instructions'. However, in the Japanese text, 'your hands' is translated directly to 'あなたの手 *anata no te'*. This is actually uncommon in natural Japanese text and it is most likely caused by the influence of the English source text.

Another interesting note was that other than pronouns, both Mandarin Chinese and Japanese tend to use classifiers anaphorically. Most often classifiers which are paired with numerals, interrogatives and determiners can be used as anaphors such as '那间 *na4jian1* (that+classifier)' which can mean 'that house/room'. Without the need of the proper noun in Mandarin Chinese, the determiner+classifier word can be used to refer to a certain room, thus acting like a pronoun. Although classifiers are not as widely used in English as in Mandarin Chinese and Japanese, numerals in English can sometimes take on anaphoric roles as well.

5.5 Limitations

As choice of words is largely affected by pragmatics, the use of pronouns can be vague at times and this proved it difficult to confirm whether the pronouns were used as referential pronouns. Particularly in the English source text, there were a few cases of ambiguous pronouns and it was difficult to confirm them as a real pronoun. For example in (5j),

5j) English: ***They*** say that away down in the village, and even in the distant parsonage, that cry raised the sleepers from their beds.

The pronoun 'they' in the sentence seem to refer to a particular group of people, however, it is actually not that simple. 'They' over here do not have a specific or tangible referent, instead it refers to something more general like the general public.

Another limitation to the study was that the corpora were segmented by POS and some words were found to be segmented wrongly, leading to us not being able to tag the pronoun. Following is an example (5k):

5k) English: ***You*** see that we have been as good as our word
Chinese: ***你瞧*** , 我们 是 说 到 做 到 的
Ni3qiao2, wo3men shi4 shuo1 dao4 zuo4 dao4 de4
'You see, we do what we say'

In the Mandarin Chinese translation, '你瞧 *ni3qiao2*' corresponds to 'You see' in the English source text. However, due to the segmentation error, '你瞧 *ni3qiao2'* was segmented as a word and this disallows us to tag '你 *ni3*' as a pronoun that can be linked to 'you' in the English text.

Also, due to the POS segmentation, semantics was not a point of consideration and sometimes this causes pronouns to be either tagged incorrectly or not be detected and thus not tagged. This problem can be solved with manual checking. However, it proves to be a challenge for the auto-tagging programme to detect pronouns through semantic use.

CHAPTER SIX

6. CONCLUSION

A qualitative and quantitative approach was used in this research to study pronouns in parallel corpora across English, Mandarin Chinese and Japanese. The results show that pronouns, though universal, are used differently across languages, resulting in a difference in distribution among the three languages and a difference in the concept links between the English-Chinese corpus and English-Japanese corpus. We have then attempted to account for these differences and presented examples of some interesting cases.

However, the pronouns in this study are solely extracted from only one novel and thus we cannot generalize the results across genres. Future work can compare other genres of texts such as non-fiction articles for their pronoun usage to see if the same results can be produced. As stated in the methodology, preliminary work has already been done on other genres of text such as an essay and on-line articles about Singapore tourism. The results may differ for different genres of text. Additionally, texts with Mandarin Chinese and Japanese as their source language can be looked at as well to control for translationese.

Also, although pronouns are a closed class of words, many of them double up as determiners. Other than the common determiners like '*the*' and '*a/an*', possessive, demonstrative and indefinite pronouns can function as determiners as well. For example, ***my*** *car* (possessive reference), ***this*** *car* (demonstrative reference), ***some*** *people* (indefinite reference). Future research can go into the study of these determiners with their relation to pronouns and how they are used across the three languages.

With this study, we hope that translation issues regarding pronoun usage would be useful and clearer to those who are learning the language and that the material from this study can contribute to pronoun translation across languages.

REFERENCES

Backhouse, A. E. (1993). *The Japanese language: An introduction*. Melbourne: Oxford University Press.

Baker, M. 1992. *In Other Words: A Coursebook on Translation*. London and New York: Rout- ledge.

Balogh, J. E. (2003). *Pronouns, prosody, and the discourse anaphora weighting approach.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI No. 3112198)

Bond, F., & Wang, S. (2014). *Issues in Building Sense-tagged Parallel Corpora with Wordnets*. Proceedings of The 7th Conference on Global WordNet. Tartu,Estonia. (to appear)

Bond, F., Wang, S., Gao, H., Mok, S., & Tan, Y. (2013). *Developing Parallel Sense-tagged Corpora with Wordnets*. Proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Discourse, Workshop of The 51st Annual Meeting of the Association for Computational Linguistics (ACL-51). Sofia, Bulgaria. 149-158.

Borthen, K. (2010). On how we interpret plural pronouns. *Journal of Pragmatics, 42*, 1799-1815.

Butt, M., Dyvik, H., King, T. H., Masuichi, H., & Rohrer, C. (2002). *The Parallel Grammar Project.* In Proceedings of COLING-2002: Workshop on Grammar Engineering and Evaluation, Taipei, Taiwan.

Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English: A comprehensive guide spoken and written English grammar and usage*. Cambridge, UK: Cambridge University Press.

Chen, J., & Wu, Y. (2011). Less well-behaved pronouns: Singular *they* in English and plural *ta* 'it/he/she' in Chinese. *Journal of Pragmatics, 43*, 407-410.

Collins Cobuild English grammar. (2005). Glasgow: HarperCollins.

Coussé, E., & Auwera, J. (2012). Human impersonal pronouns in Swedish and Dutch: A contrastive study of *man* and *men*. *Languages in Contrast, 12*(2), 121-138.

Ishiyama, O. (2008). *Diachronic perspectives on personal pronouns in Japanese.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI No. 3307632)

Kashima, E. S., & Kashima, Y. (1998). Culture and language: The case of cultural dimensions and personal pronoun use. *Journal of Cross-Cultural Psychology, 29*(3), 461-486.

Kim, C. K. (2009). Personal pronouns in English and Korean texts: A corpus-based study in terms of textual interaction. *Journal of Pragmatics, 41*, 2086-2099.

Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. California: University of California Press.

Mair, C. (2009). [Review of the book *Agreement patterns in English: Diachronic corpus studies on common-number pronouns,* by M. Laitinen] *Folia Linguistica, 43*(2), 499-502.

Ng, J. (2011). *(In)alienable possession in Chinese contrasted with English.* (Unpublished Bachelor Final Year Project).

Obana, Y. (2000). *Understanding Japanese: A handbook for learners and teachers*. Tokyo: Kurusio Publishers.

Ono, T., & Thompson, S. A. (2003). Japanese (w)atashi/ore/boku 'I': They're not just pronouns*. *Cognitive Linguistics, 14*(4), 321-347.

Ranta, A. (2009). The GF Resource Grammar Library. *Linguistic Issues in Language Techology, 2*(2), 1-63.

Ross, C., & Ma, J. S. (2006). *Modern Mandarin Chinese grammar: A practical guide*. Oxon: Routledge.

Smith, K. (2004). 'I am me, but who are you and what are we?': The translation of personal pronouns and possessive determiners in advertising texts. *Multilingua, 23,* 283-303.

Sun, C. (2006). *Chinese: A Linguistic Introduction*. Cambridge University Press

Viola, T. (2011). Philosophy and the second person: Peirce, humboldt, benveniste, and personal pronouns as universals of communication. *Transactions of the Charles S. Peirce Society, 47*(4), 389-419.

Wong, M. L., (2010). "There are many ways to translate it": Existential constructions in English-Chinese translation. *Languages in Contrast, 10*(1), 29-53.

Yip, P. & Rimmington, D. (1997). *Chinese: An essential grammar*. London: Routledge.