

The Features of Translated Language in Children's Literature Translated from English to Chinese

A thesis submitted in fulfilment of the requirements for the degree of Master of Research in
Linguistics (Translation Studies) at Macquarie University, Sydney, Australia

By

Xiaomin Zhang

Department of Linguistics

Faculty of Human Sciences

Macquarie University

October 2017

Table of Contents

List of tables	iv
List of figures	v
Abstract	vi
Statement of candidate	vii
Acknowledgements	viii
Chapter 1: Introduction	1
1.1 Introduction and rationale for the study	1
1.2 Aims of the study	3
1.3 Research questions	3
1.4 Methodology	3
1.5 Research overview	4
Chapter 2: The recurrent features of translated language and children’s literature in Chinese	5
2.1 Introduction	5
2.2 The features of translated language.....	5
2.2.1 Explication	6
2.2.2 Simplification	8
2.2.3 Normalisation	10
2.2.4 Transfer, interference or “shining through”	11
2.3 The features of translated language in translation from English to Chinese.....	12
2.4 The features of translated language and text type	15
2.5 Children’s literature and the features of translated language	16
2.6 Translated Chinese children’s literature and the features of translated language	18
2.7 Conclusion.....	20
Chapter 3: Methodology	22
3.1 Introduction	22
3.2 Research questions	22

3.3 The Comparable Corpus of Translated and Non-translated Chinese Children's Literature: Corpus composition, text collection, and text processing	23
3.3.1 Corpus composition	23
3.3.2 Corpus compilation, processing and segmentation	25
3.4 Features, indicators and operationalisations selected for investigation	26
3.5 Data collection.....	27
3.5.1 Frequency of conjunctions	27
3.5.2 Frequency of personal pronouns.....	31
3.5.3 Standardised type/token ratio (STTR)	33
3.5.4 Mean sentence length	34
3.5.5 Frequency of modal particles	34
3.6 Data analysis	37
Chapter 4: Findings and discussion	39
4.1 Introduction	39
4.2 Findings and discussion	39
4.2.1 Conjunctions.....	39
4.2.2 Personal pronouns	45
4.2.3 Standardised type-token ratio (STTR) and mean sentence length	57
4.2.4 Modal particles	59
4.3 Summary and conclusion	66
Chapter 5: Conclusion.....	70
5.1 Introduction	70
5.2 Review of the methodology	70
5.3 Summary of findings.....	71
5.4 Implications, limitations and future studies	72
References.....	75
Appendices.....	85

List of tables

Table 3.1 : Features selected for investigation.....	27
Table 3.2: List of ten most frequent potentially conjunctive tokens in the combined corpus .	29
Table 3.3: Selected conjunctions for investigation	31
Table 3.4: Personal pronouns in Chinese.....	33
Table 3.5: Selected modal particles for investigation.....	36
Table 3.6: Mood type classification.....	37
Table 3.7: Classification of modal particles by mood type in modern Chinese	37
Table 4.1: Conjunctions demonstrating no significant difference in frequency	41
Table 4.2: Pronouns demonstrating no significant difference in frequency	47
Table 4.3: Pronouns demonstrating a significant difference in frequency	47
Table 4.4: Modal particles demonstrating no significant difference in frequency	60

List of figures

Figure 3.1: A fragment of segmented paragraph in the NCCLC	26
Figure 4.1: Normalised frequency of all conjunctions (per 1,000 words) by translated status	40
Figure 4.2: Normalised frequency of 因为 <i>yīnwèi</i> ‘because’ (per 1,000 words) by translated status	41
Figure 4.3: Normalised frequency of all pronouns (per 1,000 words) by translated status	46
Figure 4.4: Normalised frequency of 咱们 <i>zánmen</i> ‘we’ (per 1,000 words) by translated status	55
Figure 4.5: STTR by translated status.....	57
Figure 4.6: Mean sentence length (in characters) by translated status	58
Figure 4.7: Normalised frequency of all modal particles (per 1,000 words) by translated status	59
Figure 4.8: Normalised frequency of 呢 <i>ne</i> (per 1,000 words) by translated status	60
Figure 4.9: Normalised frequency of 呀 <i>ya</i> (per 1,000 words) by translated status	61

Abstract

Much research in Descriptive Translation Studies has focused on the proposal that translated language demonstrates unique linguistic features, compared to non-translated language. Most studies of these recurrent features of translated language depart from Baker's (1993) formulation of four translation universals: simplification, explicitation, normalisation and levelling out. The claim that these features are "linked to the nature of the translation process rather than the confrontation of specific linguistic systems" (Baker, 1993, p. 243) points to the idea that the features of translated language are supposed to be the result of the translation process in itself, existing regardless of text type, language pair or context involved. In addition to these four features, interference, transfer or "shining through" effects are also often identified as a recurrent feature of translated language.

Against this background, this study investigates the features of Chinese translated from English in a specialised corpus of children's literature. It may be proposed that the features of translated language would be particularly salient in translated children's books, as a consequence of the importance assigned to the needs of the child reader. The study investigates simplification, explicitation and normalisation in a self-built comparable corpus of translated and non-translated Chinese children's books. In addition to these features, it also considers "shining through" of the source language as a possible feature. The objective of this study is to determine whether translated children's literature demonstrates the features mentioned above.

To answer this research question, a set of linguistic operationalisations of explicitness, complexity and conventionality were investigated. The independent samples *t*-test or Mann-Whitney *U*-test was used to determine whether these linguistic operationalisations demonstrate significant differences in the translated Chinese children's books compared to the non-translated Chinese children's books. In addition to the quantitative analysis, the study includes qualitative analysis of particularly conjunction use, optional subject pronouns and modal particles in order better to understand the quantitative findings.

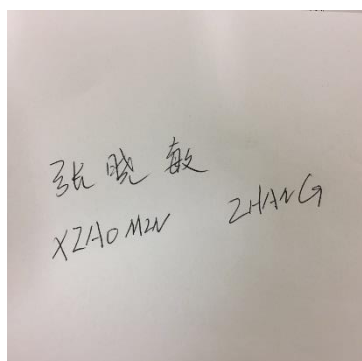
Keywords: features of translated language, English-Chinese translation, children's literature, corpus-based approach

Statement of candidate

I certify that the work in this thesis entitled “The features of translated language in children’s literature translated from English to Chinese” has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree to any other university or institution other than Macquarie University.

I also certify that the thesis is an original piece of research and it has been written by me. Any help and assistance that I have been received in my research work and the preparation of the thesis itself have been appropriately acknowledged.

In addition, I certify that all information sources and literature used are indicated in the thesis.

A photograph of a piece of paper with handwritten text. The text includes the Chinese characters '张晓敏' (Zhang Xiaomin), the pinyin 'XIAOMIN', and the English name 'ZHANG'.

Name: Xiaomin Zhang

MQ ID: 45002959

9 October 2017

Acknowledgements

I would first like to thank my supervisor, Dr Haidee Kruger, for her continuous support, inspiration, encouragement and patience throughout my study. Her enlightening guidance helped me at every stage of this study and when writing this thesis. Without her, this thesis would not have been possible.

I also would like to thank my co-supervisor, Dr Jing Fang, for her valuable comments and suggestions, and helping me with her immense knowledge of functional grammar.

My very profound gratitude goes to my parents for their unconditional love and support throughout this study. Their unfailing support and continuous encouragement have always been my source of hope and success.

Chapter 1: Introduction

1.1 Introduction and rationale for the study

The concept of “recurrent features of translated language” derives from Baker’s (1993) proposal that translated language is qualitatively different from non-translated language, and that it has unique linguistic features which distinguish it from non-translated language. The origins of this idea may be seen to date back to Frawley’s (1984) notion of “a third code” and Toury’s (1995) proposal of “laws of translation”. Various terms have been used to refer to the features of translated language, including “translationese” (Gellerstam, 1986), “universal features of translation” (Baker, 1993), “translation universals” (Mauranen & Kujamäki, 2004), “mediation universals” (Ulrich & Murphy, 2008), “features of translation” (Olohan, 2004), and “recurrent features of translated language” (Kruger & Van Rooy, 2012; Redelinghuys & Kruger, 2015).

These significant differences are proposed as “independent of the influence of the specific language pairs involved in the process of translation” (Baker, 1993, p. 243) and are ascribed to some effect that is “inherent in the translation process itself” (Baker, 1993, p. 246). In other words, the recurrent features of translated language are assumed to be a “universal” quality of translations, existing regardless of text type, language pair or context involved. The proposed features include Baker’s (1996) formulation of four translation universals: explicitation, simplification, normalisation and levelling out (Baker, 1993; Laviosa, 2008).

However, the notion of “universals” has been widely questioned, as have some of the early assumptions in this area of research. Some scholars argue these features are not specific to the practice of translation but are common to all communication taking place in situations of contact (House, 2008; Kruger & Van Rooy, 2016; Lanstyák & Heltai, 2012). Moreover, there are questions about whether the “universals” of translated language transcend text type differences (Kruger & Van Rooy, 2012, 2016), as well as differences in translation processes, for example, in manual, machine and machine-assisted translation (Lapshinova-Koltunski, 2015). In addition, many researchers have questioned the initial exclusion of source-text and source-language influence, and identify the effects of source-language transfer, interference or “shining through” as another recurrent feature (Teich, 2003; Toury, 2012).

Subsequent to the earlier studies carried out by Baker and colleagues, which were primarily focused on the “universals” formulated by Baker (1993), other features have also subsequently been proposed and investigated. These include transfer- or interference-related features (Hansen-Schirra, 2011; Mauraanen, 2004; Teich, 2003); the translation of unique items (Cappelle, 2012; Eskola, 2004) following the Unique Items Hypothesis of Tirkkonen-Condit (2002); and asymmetry between implicitation and explicitation (Becher, 2010; Klaudy & Karoly, 2005). However, comparatively greater attention has been given to the features originally proposed by Baker (1993). While Baker’s original proposal of these features as “universals” has subsequently been revised to rather view these features as conditioned and probabilistic tendencies of translation (Toury, 2004), there is general agreement that translations tend to demonstrate increased explicitness, decreased complexity, increased conventionality and reduced variability, when compared to non-translations (see Zanettin, 2013).

Early research on the features of translated language focused on translated English, and made use of corpora of relatively restricted text types (see the detailed discussion in Section 2.2). Over the past two decades, however, there have been considerable expansion of research in this area. A wide range of languages have been studied, with a particular expansion in studies of translated Chinese (see Xiao, 2010, 2011; Xiao & Hu, 2015, and the further discussion in Section 2.3). While there has been a growing awareness of the important ways in which text type or register condition the realisation of the features of translated language (see Kruger & Van Rooy, 2012, and the further discussion in Section 2.4), studies of the features of translated language in different text types represent a relatively new area of research. An important point made by researchers working on register-based studies of the features of translated language is that the particular aims and features of a text type or register may probabilistically condition the realisation of the features of translated language.

In this study, I propose that translated literary works for children may be a potentially fruitful text type in which to investigate the features of translated language, since the audience, nature and aims of children’s books may predispose translators towards increased explicitation, simplification and normalisation, all in aid of meeting the perceived needs and expectations of the target audience of child readers (see Section 2.5 for more detailed discussion).

1.2 Aims of the study

Despite the potential that translated children's literature holds as a unique register in which to investigate the features of translated language, there has been very limited research in this area. This is particularly the case for translated Chinese children's literature: there have been very few comprehensive and systematic corpus-based investigations of the features of translated language in this text type (see Section 2.6). This study addresses this gap by investigating the features of translated language in children's literature translated from English to Chinese, utilising a combination of quantitative and qualitative corpus-based methods.

1.3 Research questions

Against the background of the literature review in Chapter 2, this study aims to answer the following questions:

1. Does Chinese children's literature translated from English demonstrate evidence of the following three proposed recurrent features of translated language, in comparison with non-translated Chinese children's literature?
 - a. Increased explicitness
 - b. Simplification
 - c. Normalisation
2. What is the likely motivation for any differences in explicitness, complexity and conventionality observed between Chinese children's literature translated from English, in comparison with non-translated Chinese children's literature? In particular, is there evidence that source-text transfer, interference or "shining through" may account for observed differences?

1.4 Methodology

In order to answer research question 1, a quantitative corpus-based method was utilised. A monolingual comparable corpus consisting of Chinese children's literature translated from English and non-translated original children's literature in Chinese was compiled. These two subcorpora were constructed to be as comparable as possible in size, text type, historical timeframe of production and other relevant parameters. Five linguistic features were selected as operationalisations for the three recurrent features of translated language investigated in this study. Conjunction and pronoun use was investigated as indicative of explicitation.

Standardised type-token ratio and mean sentence length were used as operationalisations of simplification. Lastly, the use of modal particles was analysed as a measure of normalisation. Data collection was carried out by using various functions in WordSmith Tools 7.0 (Scott, 2016). The two-sample *t*-test and its non-parametric alternative, the Mann-Whitney *U*-test, were used to assess whether the two subcorpora demonstrated significant differences for each of the five features investigated.

Research question 2 was answered by qualitative analysis of the use of particular conjunctions, optional personal pronouns, and modal particles, in order to explore possible reasons for the observed differences in explicitness, complexity and conventionality. While the comparable corpus design of this study does not allow for a quantitative assessment of the potential effects of source-language transfer or interference, qualitative comparisons of the translations and their source texts are also used to explore this possibility.

1.5 Research overview

Following this introductory chapter, Chapter 2 presents a comprehensive literature review on the features of translated language, proposed explanations for these features, and their realisation in relation to language pairs and text type, with a particular emphasis on Chinese children's literature. This discussion paves the way for the investigation of the features of translated language in children's literature in China and also provides the necessary background to the selection of the operationalisations used in the present study. Chapter 3 outlines the methodology used in this study. It describes the details of the corpus composition and compilation; the operationalisations investigated in relation to explicitation, simplification and normalisation; and the data collection and analysis. The findings and discussion are presented in Chapter 4, where the quantitative and qualitative findings for each operationalisation are reported and discussed in turn. Chapter 5 summarises the main findings and conclusions of this study, and outlines the limitations of the study as well as future avenues of research.

Chapter 2: The recurrent features of translated language and children's literature in Chinese

2.1 Introduction

This chapter consists of five sections, each focusing on key conceptual and theoretical issues that form the necessary background to this study. Section 2.2 provides a general overview of the features of translated language, focusing on definitions and explanations offered for these features as well on some influential findings. The two subsequent sections highlight important factors that condition the realisation of the features of translated language. The effect of different language pairs in translation is considered in Section 2.3, with particular attention to studies of the features of translated language involving English-Chinese translation, the focus of this study. In Section 2.4, the relationship between the features of translated language and text type is discussed. The two concluding sections draw together the earlier discussions, and focus on the relation between the features of translated language and children's literature, specifically in China. Section 2.5 explains the rationale for and importance of investigating the features of translated language in the text type of children's literature, and briefly discusses some studies on the features of translated language in children's literature in European languages. Section 2.6 narrows the focus even further, considering existing research on translated Chinese children's literature and specifically studies of the features of translated language in this text type in Chinese. This section concludes by identifying the research gap that this study aims to address.

2.2 The features of translated language

As outlined in Section 1.1, the idea that translated language demonstrates several unique linguistic features that distinguish it from non-translated language has motivated the search for recurrent or typical features of translation. These are often defined as features which “typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems” (Baker, 1993, p. 243). However, some scholars also see source-language interference features or source language “shining through” (Teich, 2003) as characteristic of translated language. This is reflected in Chesterman's (2004) conceptual distinction between S-universals and T-universals (Chesterman, 2004). T-universals refer to differences between translations and non-translations in the same language (the target language), investigated by means of comparable corpora; while S-universals refer

to both similarities and differences between translations and their source texts, investigated by means of parallel corpora (Chesterman, 2004, pp. 8). In this section I briefly discuss four features of translated language that form the focus of this investigation: explicitation, simplification, normalisation, and “shining through” or transfer effects.¹

2.2.1 Explicitation

Explicitation refers to the tendency for translations to be more explicit in lexicogrammatical encoding, compared to both their source texts (explicitation as an S-universal) and to non-translated texts in the target language (explicitation as a T-universal). Explicitation is often regarded as the consequence of cognitive effort in translation. The process of translating is seen as cognitively effortful, since translators are “shuttling” between two languages both simultaneously and continuously (Kruger & Van Rooy, 2016). Following on the “complexity principle” proposed by Rohdenburg (1996, p. 149), which posits that “more explicit grammatical alternatives tend to be preferred in cognitively more complex environments” it is argued that increased explicitness is therefore a cognitive “crutch” for translators, easing processing difficulty (Kruger & Van Rooy, 2016, p. 29).

Other explanations are more socio-cognitive in nature, and are particularly focused on how translators construe the needs of the reader. As proposed by Saldanha (2008), the increased explicitness in translated texts may be the result of translators’ awareness of their communicative roles in relation to readers and assumptions about their readership. Translators try to provide more “communicative clues” to help their readers, who do not share as much cultural ground with the author as the readers of the source text do (Pym, 2005). Lastly, explicitation may be the consequence of risk avoidance on the part of translators. Translators, as Pym (2005) argues, are likely to avoid risks that could affect the communication between participants involved, because the risk of non-cooperation in communication could lead to translators losing income or the trust of their clients (Pym, 2005, p. 34). This risk management may account for the feature of explicitation.

In terms of S-universal explicitation (also referred to as “S-explicitation” by Krüger, 2014), implicit information in the source text is spelled out by adding extra elements in the translation, for example providing supplementary explanatory material, expanding condensed passages and clarifying ambiguities (Olohan & Baker, 2000). Several studies, using parallel corpora and

¹ The feature of levelling out is not investigated in this study, and is therefore not discussed in further detail.

often investigating the relationship of explicitation and implicitation in translation, have found support for S-explicitation of this kind (see Blum-Kulka, 1986; Kenny, 2004; Øverås, 1998; Vanderauwera, 1985).

T-universal explicitation can be operationalised at linguistic levels of syntax and discourse, though Puurtinen (2004) also analyses lexical explicitation in Finnish children's literature (see Section 2.5). At the level of syntax, optional syntactic elements such as the optional complementiser *that* (Olohan & Baker, 2000) and optional personal pronouns and articles (Jiménez-Crespo, 2011) have been used as an indicator of syntactic explicitation. Despite corpus composition differences in translation modalities and types, studies like these have provided evidence for the assumption of increased explicitness of lexicogrammatical encoding in translated texts.

At the level of discourse, indicators of explicitation may include the increased use of explicative reformulation (Xiao, 2011). Generally, the more frequent appearance of reformulation markers in translated Chinese and translated English in contrast to their native counterparts supports the observation that reformulation markers function as a strategy to increase explicitness in translations (Xiao, 2011). Other discourse features investigated include conjunctions and discourse particles. For example, Pápai (2004) shows that translation-related explicitation is evident in the translation process from English into Hungarian, marked by a higher frequency of conjunctions and discourse markers in translated Hungarian. (See Section 2.3 for more detailed discussion of other explicitation features, specifically related to translated Chinese.)

While explicitation is one of the most frequently investigated features of translated language, the explicitation hypothesis is not uncontroversial. Becher (2010) is one critical voice, stating that “the dogma of translation-inherent explicitation rests on fallacious theoretical considerations and premature interpretations of empirical data” (Becher, 2010, p. 1). Olohan and Baker (2000) interpret their findings of a higher incidence of the optional *that* in translated texts as evidence of “inherent, subliminal processes of explicitation in translation” (Olohan & Baker, 2000, p. 143). Becher (2010) offers alternate explanations for the higher frequency of the complementiser *that* in translated texts, including source-language interference (Saldanha, 2008) and translators' conservatism in preferring to use a more formal style (Swan, 1980). He argues strongly in favour of “abandoning the notion of ‘translation-inherent’ explicitation” and replacing it with the asymmetry hypothesis to interpret explicating and implicating shifts, as

in Øverås's (1998) study, since different languages have different lexicogrammatical and stylistic preferences in respect of the degree of explicitness (Becher, 2010). The asymmetry hypothesis posits that translation-inherent explicitation can only be proven by investigating bi-directional translation, when explicitation takes place in the source language → target language direction, but implicitation is not observed in the target language → source language direction. This phenomenon is referred to as asymmetric explicitation (Klaudy & Karoly, 2005, p. 13).

The cause of explicitation and implication, according to Klaudy and Karoly (2005), can be found both in “language-specific” and “non-language-specific” factors. Klaudy (2008) further distinguishes between different types of explicitation based on the motivation for their occurrence.² Apart from “translation-inherent explicitation”, obligatory explicitation is caused by the lexicogrammatical differences between the two languages involved, which means “if no explicitation, the TT sentence will be ill-formed” whereas optional explicitation is generally the result of different stylistic preferences between source language and target language (Klaudy, 2008, pp. 106-107).

2.2.2 Simplification

A further challenge to explicitation is the fact that it could be treated as a manifestation of simplification, as a more explicit style also correlates with a simpler style (Xiao, 2011; Zanettin, 2013). Baker (1996) explains that “simplification involves making things easier for the reader (but not necessarily more explicit), but it does tend to involve also selecting an interpretation and blocking other interpretations, and in this sense it raises the level of explicitness by resolving ambiguity” (Baker, 1996, p. 182). This means that the indicators of explicitation and simplification might not always be that easy to distinguish.

Simplification refers to the “tendency to simplify the language used in translation” (Baker, 1996, p. 181). As is the case for explicitation, both cognitive and socio-cognitive explanations have been offered for the phenomenon. Baker's (1996) definition of simplification set the ground for assuming that simplification is the consequence of translators' prioritisation of readers' needs: “Simplification involves making things easier for the reader” (Baker, 1996, p. 182). The lack of similar lexical structures and/or comparable expression of cultural concepts might compel translators to fill this gap for their readers, which could potentially result in lexical simplification (Laviosa, 2002). Halverson (2003) focuses on a more restricted

² Due to limitations of length, the various types of explicitation are not discussed in further detail here. See Klaudy (2008) for an overview.

psycholinguistic framework to account for the features of decreased complexity, from the perspective of cognitive linguistics. According to Halverson (2003), the gravitational pull of highly salient nodes and structures (the highest-level schema) in the target language may result in a narrower range of lexical choices made by translators, and these prototypes, once established, are more frequently used in translation (Halverson, 2003). As is the case for increased explicitness, cognitive effort may constrain translators' lexical choices towards more frequently used vocabularies since the more infrequent, diverse options might not be easily accessible during cognitively demanding language processing (Kruger & Van Rooy, 2016).

Considering translation as an S-universal, Malmkjær (1997) finds stylistic simplification in translations, reflected in the tendency of translators to alter weaker punctuation to stronger; for example, commas to semicolons or full stops, semicolons to full stops. As a consequence, long and complex sentences in the source text are turned into shorter and simpler ones in the target text (Xiao, 2010).

T-universal simplification can be observed at lexical, syntactic and stylistic levels. Lexical simplification, as manifested by lexical variety and lexical density, has been investigated by a number of scholars, such as Corpas Pastor, Afzal, and Pekar (2008), Cvrček and Chlumská (2015), Kruger and Van Rooy (2012), Laviosa-Braithwaite (1997), Laviosa (1998) and Xiao (2010). Lexical variety can be measured by type-token ratio (the ratio of the number of different words to the number of running words in a text), which reflects vocabulary range. Lexical density can be measured by the ratio of the number of content items, as opposed to grammatical items,³ to the number of running words in a text, thus reflecting information load (Puurтинен, 2003). Using measures such as these, Laviosa's (1998) studies of newspaper and narrative translated texts in English yield supporting evidence for simplification. However, other studies have not found support for simplification at the lexical level (e.g. Jantunen, 2001).

Indicators of syntactic simplification might include more use of finite than non-finite constructions (Puurтинен, 2003). Non-finite constructions can contain a large amount of information in a compact form, and are associated with an implicit style characterised by the absence of connectives, which could lead to decreased readability and speakability (Puurтинен, 2003). However, Puurтинен's (2003) study of translated and non-translated Finnish provides

³ "Content items" is used in line with Halliday and Matthiessen (2004), referring to the names of entities, processes and qualities, for example, *water*, *move*, *beautiful*; "grammatical items" or "function words" (Hu, 2007; Xiao, 2010; Xiao & Hu, 2015) refer to "words functioning as the direct realisation of terms in grammatical systems", for example *the*, *on* and *if* (Halliday & Matthiessen, 2004, p. 44).

evidence that contradict the hypothesis of increased simplification in translation: non-finite constructions appear more frequently in translated texts (Puurtilinen, 2003).

Overall, there is mixed support for the simplification hypothesis. In particular, studies that use a combination of measures (e.g. Corpas Pastor *et al.*, 2008) highlight diverging trends in relation to simplification and complexification. (See Section 2.3 for more detailed discussion of other simplification features, specifically related to translated Chinese, where similar conflicting trends are identified.)

2.2.3 Normalisation

A strong preference for conformity to conventions or norms in the target language (TL) is referred to as normalisation (also known as “conventionalisation,” “standardisation,” or “conservatism”), sometimes even to the extent of exaggeration (Baker, 1996, p. 177). It is related to Toury’s law of growing standardisation, which posits that “in translation, textual relations obtaining in the original are often modified, sometimes to the point of being totally ignored, in favour of [more] habitual options offered by a target repertoire (Toury, 2012, p. 304).

Cognitive explanations for normalisation have been offered. The same cognitive mechanisms accounting for simplification could also result in normalisation. In other words, the category prototype and highest-level schema of the target language might also exert a gravitational pull, which encourages the occurrence of “specific TL lexical and grammatical structures that correspond to those salient nodes and configurations in the schematic network” (Halverson, 2003, p. 218). As a consequence of these, an overrepresentation or exaggeration of specific target-language lexical and grammatical features occurs in translation, which accounts for the tendency of normalisation (Halverson, 2003, pp. 218-221). In contrast to this cognitive view, however, normalisation is most often regarded as a consequence of socio-cultural or economic constraints (Kenny, 2001). Because translators translate with readers’ expectations about acceptability in mind, this biases translators to translate towards existing norms and conventions. Translations that deviate from these expectations for acceptability might cause the translation to be criticised, ignored and rejected by the target audience (Kenny, 2001, p. 67).

Normalisation can be viewed as both an S- and a T-universal as translations could be more conventional, more conservative and less creative compared to their source texts as well as to non-translated texts in the target language. Investigations of normalisation have frequently

focused on lexis. Kenny (2001), for example, exemplifies the investigation of normalisation as an S-universal, at the lexical level. In terms of T-universal normalisation, lexical normalisation may be reflected in a lower degree of lexical productivity (Olohan, 2004) and a higher frequency of conventional lexical patterns or collocations (Baker, 2007; Mauranen, 2000). However, while there is evidence that translators do overuse conventional lexical patterns, Mauranen (2000) also shows that translations tend to reflect more “unusual word combinations” than originals, which provides evidence against normalisation (Mauranen, 2000, p. 120, 137). Mauranen (2000) proposes that these “strange strings” could be the consequence of interference from source-language collocational preferences. For normalisation, too, findings from existing studies provide mixed support (see also Section 2.3 for discussion on these features in studies of Chinese translation), and there appears to be a particularly strong tension between normalisation and transfer effects.

2.2.4 Transfer, interference or “shining through”

According to Toury’s (2012) law of interference, “phenomena pertaining to the make-up of the source text tend to force themselves on the translators and be transferred to the target text” (Toury, 2012, p. 275). Interference is also termed transfer or “source-language shining through” (Teich, 2003) and refers to the idea that translations show influence from the source language (SL) or source text, due to the basic cognitive condition of the prior cognitive activation of the source language, leading to what has been described as cross-linguistic priming effects (Kruger & Van Rooy, 2016). It has been widely investigated, at both the lexical level (e.g. Mauranen, 2004) and the syntactic level (e.g. Hansen-Schirra, 2011). For example, Mauranen (2004) compares the relative distance between translated Finnish corpora from individual source languages, a mixed-translation Finnish corpus including several source languages, and a non-translated Finnish corpus, by investigating the distribution of frequent items. The results show that translated texts clearly deviate from non-translated texts, and their profiles of deviation correlate to source-language effects (Mauranen, 2004, p. 79). The fact that translations resemble each other by bearing “a closer affinity to each other rather than to untranslated texts” further suggests that interference might be a universal feature of translations (Mauranen, 2004, p. 79).

The tension between normalisation and interference (shining-through) has been a focus in the work of Hansen-Schirra (2011), who investigates typical and atypical features associated with fictional writing in translated and non-translated English texts (using the TEC, BNC and the

CroCo corpus).⁴ The findings show that both typical and atypical features occur more frequently in the TEC compared to the BNC, meaning that translations show contradictory tendencies in relation to normalisation (Hansen-Schirra, 2011). Based on an analysis of the directional parallel CroCo Corpus, she demonstrates that the overuse of atypical fiction features could be ascribed to shining-through effects, due to the literal translation of the corresponding source-text structures (Hansen-Schirra, 2011, p. 147). The co-occurrence of normalisation and interference results in the hybridisation of target texts, which are dissimilar to both the source language and the target language (Hansen-Schirra, 2011).

Related to this is a special form of transfer effects termed the “Unique Items Hypothesis”, which refers to the finding that items which are unique to the target language and not present in the source language tend to be under-represented in translations compared to non-translations in the same language, because their selection is “inhibited” by their absence in the source language, and they therefore occur less frequently (Tirkkonen-Condit, 2002; Eskola, 2004).

Much of the focus in conceptual discussions of the features of translated language has been on questions of universality, following on Baker’s (1993) formulation of these features as “universals” of translation. The current agreement is that these features are not universals, but rather probabilistically conditioned tendencies of translated language. In other words, the realisation of these general tendencies of translated language is conditioned by a variety of factors, including the languages involved (Lefer, 2012), translation processes (Lapshinova-Koltunski, 2015) and text type (Kruger & Van Rooy, 2012). Two of these factors (text type and language pair) are of particular interest in this study, with its focus on the translation of **children’s books** from **English to Chinese**. The following section considers first the matter of language-specificity, focusing specifically on some studies of the features of translated language involving Chinese.

2.3 The features of translated language in translation from English to Chinese

Work on translated English dominated early research on the features of translated language (see Baker 1996, 2007; Laviosa, 1997, 1998; Olohan & Baker, 2000; Olohan, 2004). Over time, however, the range of languages studied has widened and more language pairs have been

⁴ “TEC” stands for Translational English Corpus while “BNC” stands for British National Corpus. The CroCo Corpus is a bidirectional parallel corpus, smaller than the TEC, consisting of English originals and their German translations as well as German originals and their English translations (Hansen-Schirra, 2011, p. 143)

involved. Initial expansion involved mostly European languages, but more recently, Chinese has formed a strong focus of investigation.

Within the framework of corpus-based investigations of the features of translated Chinese, explicitation has frequently been discussed at lexical and/or syntactic levels, with some focus on simplification as well (e.g. Hu, 2006; Huang 2007, 2008; Ke 2005; Wang & Qin 2010, Xiao, 2010, 2011; Xiao & Hu, 2015). Xiao (2010) along with others (Hu, 2006; Hu & Zeng, 2009; Huang, 2007; Ke, 2005) have investigated the use of conjunctions in Chinese translation from English. They all find a higher frequency of conjunctions in translated texts than in native non-translated texts, providing support for the explicitation hypothesis. Huang (2007), Wang and Hu (2010) and Xiao and Hu (2015) compare the frequency of personal pronouns in translated and non-translated texts in Chinese and find that pronouns show significantly higher frequency in translations. Moreover, Wang and Hu (2010) find the frequency of re-occurrence of the third person pronoun 他 *ta* ‘he’ as well as its anaphoric function has been noticeably strengthened in translations. These studies furthermore suggest that the more frequent use of personal pronouns may be influenced by the source language, English. Average sentence length in translated Chinese texts tend to be longer than in non-translated texts in studies conducted by Hu (2006) and Wang and Qin (2010). On the one hand, this could be taken as evidence for explicitation, as translated Chinese prefers to convey the meaning more clearly by adding more words. On the other hand, it could be taken as evidence against simplification, which is commonly associated with shorter average sentence length (see Section 2.2.2). Wang and Qin (2010), too, explain that longer sentence length might be induced by the source language, for example, through the translation of *that*. Translators might add more words to explicitate the relation between main and complement clauses since there is no equivalent for *that* in Chinese.

In addition, Ke (2005) proposes coexisting patterns of explicitation and implicitation in translated language. When translating from a “highly grammatically explicit language”, which prefers to use more function words to systematically connect sentence components, like English; to a “grammatically implicit language”, which opts to use fewer function words, like Chinese, explicitation increases and implicitation decreases (Xiao & Hu, 2015, p. 28). In a reverse translation direction, the tendency of explicitation decreases and implicitation increases (Xiao & Hu, 2015). These findings suggest that explicitation is, at the very least, conditioned by preferences for explicitness of encoding in different languages, and there is therefore a transfer dimension to explicitation, as also pointed out in Section 2.2.1.

In terms of simplification, standardised type-token ratio (STTR), analysed by Hu (2006), Wang and Qin (2010), Xiao (2010), and Xiao and Hu (2015) yields contradictory results. Wang and Qin (2010) find that STTR is higher in translated texts in comparison with non-translated texts, thus contradicting the simplification hypothesis, whereas Hu (2006) finds that STTR is lower in a translated Chinese corpus, suggesting a smaller range of vocabulary in translations. Furthermore, Xiao (2010) and Xiao and Hu (2015) find no significant difference in STTR between translated and non-translated texts. The differences in findings might be explained by the factor of text type since Hu (2006) is based on fiction translations, while other studies are based on general Chinese (see Xiao & Hu, 2015 for an overview). The ratio of lexical words to function words in translations generally does demonstrate lower lexical density in translated Chinese, which supports the simplification hypothesis (Hu, 2006; Xiao, 2010; Xiao & Hu, 2015).

Xiao and Hu (2015) also investigate modal particles as a measure of normalisation and/or source-language shining through. Modal particles are of interest because they lack a corresponding equivalent in English. Comparing the frequency of modal particles in translated Chinese and native Chinese shines light on the tension between normalisation and transfer in translated language. The feature of normalisation may be seen to exist in translated Chinese when modal particles are more frequently used (and even over-represented) in the context of translated Chinese. Alternatively, transfer may be evident when modal particles are less frequently used because the activation of English limits the use of modal particles in translations due to source-language “shining through”. The findings show the use of modal particles is quite often affected by the source text and underused in translated Chinese. This points to the conclusion that source-language “shining through” may have more significant effects than target-language normalisation in translation from English to Chinese in terms of modal particles (Xiao & Hu, 2015). Further support for transfer effects come from Wang and Qin (2010), who find the suffix -性 *xing* shows a stronger word formational capacity in translation. They argue that this is the consequence of affix-by-affix translation from the English affixes *-ity*, *-ness* and *-dom*.

As pointed out above, the language pairs involved in translation is an important factor in conditioning the realisation of the features of translated language – a point illustrated by the discussion in this section. A second important factor conditioning the features of translated language is text type. The following section considers specifically the question of whether the features of translated language occur across different text types, or whether text type

significantly affects the realisation of these features. This discussion forms the necessary background to the further discussion of the text type of children's literature, and the features of translated language specifically in Chinese children's literature translated from English, presented in the subsequent sections.

2.4 The features of translated language and text type

While it is essential to be aware of the differences in language produced in different circumstances (Lee, 2001), drawing such distinctions is not always straightforward. The (related) notions of "text type", "register" and "genre" are characterised by terminological and conceptual difficulties (see Biber, 1995; Lee, 2001; Lefer, 2013). The scope of this thesis does not allow for detailed discussion of these complexities, and in this study, "text type" is therefore used in a broad sense that takes in meanings related to both genre and register, to refer to a language variety that is defined on the basis of common lexical, syntactic and discourse linguistic features that can be traced to a particular communicative context and function (Biber, 1995).

Text type is an important and meaningful entity in both contrastive linguistics and translation studies (Neumann, 2014). The significance of text type in the pursuit of the features of translated language was articulated early on: "Are certain linguistic features or strategies more likely to occur in certain types of translation genres, like translated fiction, news, inflight magazines?" (Baker, 1999, p. 292). More recently, researchers have started to consider the relationship between text type and the features of translated language in a more rigorous quantitative way.

Neumann (2014) finds evidence for text type (register) normalisation and levelling out in German texts translated from English. Translators are prone to keeping their translations close to conventional text type patterns even to the extent of over-adjusting their translations and "bleaching" text type features which are common in non-translated texts (Neumann, 2014, pp. 52-53). Kruger and Van Rooy (2012) systematically investigate the relationship between text type and the features of translated language, utilising a corpus of English texts translated from Afrikaans, and the International Corpus of English for South Africa (ICE-SA) as a comparable corpus. Their findings reveal that there is a text-type effect, in that these features are less obvious in more informational translated registers than in less formal and more creative registers.

By means of profile-based correspondence analysis, Delaere et al. (2012) aim to verify the hypothesis of normalisation in translated Belgian Dutch and to determine whether text type is an influential factor in this tendency. Based on observations of the use of standard language in the Dutch Parallel Corpus, they find translated Belgian Dutch does contain more standard language than non-translated Belgian Dutch. Furthermore, these differences are dependent on text type, as texts with more editorial work make use of more standard language compared to text types that are subject to less editorial control. In a similar vein, Lefer (2012) has demonstrated the impact of language-pair specific features and text type variation in the realisation of normalisation.

Previous studies have aimed to understand how text type may influence the realisation of the features of translated language. However, such studies are comparatively limited, which has implications for the generalisability of theories of the features of translated language. This study focuses on a text type which has not been frequently analysed in research on the features of translated language, namely children's literature. The following section will discuss the text type of children's literature in general, and argue that children's literature may be a text type in which the forces that condition the features of translated language are especially strongly felt, as part of the motivation for this study. Subsequent to this, Section 2.6 considers specifically research on translated children's literature in China, and existing studies on the features of translated language in translated Chinese children's literature.

2.5 Children's literature and the features of translated language

Children's literature has particular functions and features that may affect the realisation of the features of translated language. One possible explanation that has been proposed for the occurrence of the features of translated language (see Section 2.2) is that these features are the consequence of translators' risk management in the translation process. Translators, as Pym (2005) argues, are likely to avoid risks that could affect the communication between participants involved, because the risk of non-cooperation in communication could lead to translators losing income or the trust of their clients.

In the case of the translation of children's literature, translators may be particularly averse to risk, because there are such strong demands for target-audience acceptability in children's literature, where the target audience is both the child readers and the adults (parents, educators) who would typically read with the child, and also select and buy books (Puurтинен, 1995; O'Connell, 1999). Target-audience acceptability is of vital importance in the translation of

children's literature, as it directly influences the market and publishers' decisions on what to publish. This may affect translators' strategies and decisions (O'Connell, 1999). To attain the requisite level of acceptability, translators adjust the source text to fit in the target system, by ensuring that the translation fits existing models of the genre in the receiving system (Shavit, 2006). In this process, the text may be changed by "deleting or adding" elements to ensure that the text is adjusted to the existing models in the target system (Shavit, 2006, p. 26). Also, "undesirable scenes" (again, judged undesirable by the standards of the target system for the genre) may be deleted if the deletion will not impede understanding (Shavit, 2006, p. 35). For the translators of children's books, it is a priority to adjust "the plot, characterization and language" (Puurтинен, 2006, p. 54) to the level of children's comprehension (or at least, how this is judged in the recipient system), which often leads to shorter and less complex content. Against this background, it may be argued that translated children's literature would be particularly susceptible to demonstrating the features of explicitation, simplification and normalisation, as translators would do this to ensure that the text fit the perceived expectations of the target audience.

While there are numerous studies of translated children's literature, primarily in the context of European languages (see O'Connell, 1999; Shavit, 2006; Van Coillie & Verschueren, 2014), there are only a few studies of translated children's literature in the context of the features of translated language. As already discussed in Section 2.2.2, Puurтинен (2003) uses nonfinite constructions as a device to measure the readability of children's books, using a combined parallel and comparable corpus of children's books: English source texts, their Finnish translations, and Finnish originals. Nonfinite constructions are more complex constructions, which are also less explicit since they "pack" much information into propositions without making the relations between propositions clear (Puurтинен, 1998, p. 3). Puurтинен (2003) regards the use of nonfinite constructions as a case of translationese, which may be traced back to source-language transfer. In a subsequent study, Puurтинен (2004) investigates the frequency of clause connectives (such as conjunctions, adverbs and relative pronouns), which are used to explicate the relation (causal, temporal and post-modifying) between clauses in translated Finnish children's literature. Focusing on the question of whether translations are more explicit than non-translated originals, as evident in the more frequent use of clause connectives, she selects 13 commonly used Finnish clause connectives for investigation. The findings show that a few connectives are more frequent in translations, while others have a higher frequency in non-translated originals and some have roughly the same frequency in both. Therefore, the

findings do not fully support the explicitation hypothesis, but also do not provide clear evidence against it. Rather, explicitation appears to be conditioned by the functions of the connectives and the context of use (Puurtilinen, 2004).

This study focuses specifically on the features of children's literature translated from English to Chinese, and more background on the translation of children's books and studies of translated language in children's books in China is therefore essential. This forms the focus of the following section.

2.6 Translated Chinese children's literature and the features of translated language

In comparison with studies of translated children's literature in Europe, the study of children's literature translation in China has lagged behind and has largely been ignored by mainstream translation studies (Li, 2014). According to Li (2014), the foci of studies on the translation of Chinese children's literature have been the following aspects:

- general description of the phenomenon of children's literature translation (see Wu, 2007; Xu, 2004);
- the history of children's literature translation (see Li, 2005; Qin, 2004; Wen & Wang, 2008; Zhang, 2008);
- famous translators of children's literature (see Zhang, 2006, 2010);
- children's literature from interdisciplinary perspectives (see Hu, 2009; Song & Huang, 2010; Yuan, 2009).

After a detailed search on CNKI (the China academic journals full-text database), five masters' theses dealing with the features of translated language in children's books in China were found. Among these, Wang (2013) investigated explicitation in English-Chinese translation based on a comparable and parallel corpus. She aims to search for explicitation of logical relations, ideational meanings and emotional meanings. She studies connectives and transitional words associated with the explicitation of logical relations; the concretisation of nouns and adjectives for explicating ideational meaning; and modal particles and adverbs for emphasising emotional meaning. She finds evidence for the explicitation of logical relations, as well as ideational meaning – but not emotional meaning. Wang (2013) explains the cause of explicitation as the consequence of language differences, combined with translators' subjectivity. The major limitation of this study, as Wang (2013) points out, is the size of the corpora she used. Each corpus included only seven texts.

Zang (2010) and Mei (2015) describe translated language in children's literature at lexical and syntactic levels. Both studies use self-built corpora of translated and non-translated Chinese texts. Zang (2010) focuses on children's literature targeted at the age group of 6-12 years, while Mei's (2015) texts focus on older children, aged 12-18 years. Zang (2010) aims to explore the linguistic features of translated language with the ultimate goal to evaluate the quality of the translations and point out deficiencies. Mei's (2015) research objective is twofold. She intends to describe the linguistic features of translated Chinese children's literature in contrast with the originals. Additionally, she wants to know how these features are affected by factors of "translation universals", language differences between English and Chinese and the characteristics of children's literature. For each feature, Mei (2015) states the conditions for their appearance in relation to these three factors.

Both studies find that the Chinese translated children's books possess several distinct features compared to comparable non-translated works. These features include: a relatively higher lexical variety, a lower lexical density and a lower frequency of major content words; a less colloquial and more formal style, reflected in less frequent use of modal particles, onomatopoeia and idioms; and excessive use of pronouns and conjunctions. However, the two studies also yield some contradictory findings. For instance, Zang (2010) finds an overuse of passives while Mei (2015) finds no significant difference in the frequency of passives in her comparable corpora. Both studies provide evidence that support as well as contradict the existence of the translational features of simplification, explicitation and normalisation. However, the limitations of these two studies are apparent. Mei's (2015) study explains the influence from the above three factors mainly from an intuition-based perspective. No theoretical and practical evidence are provided. Moreover, not all the translated books she chooses for children aged 12-18 years are representative reading for the age group; for example, the translations of *Peter Pan* and *The Adventures of Tom Sawyer* she includes are actually also suitable for children under 12. Her corpus has only four texts in each subcorpus, which might limit generalisability to (translated) Chinese children's literature as a whole. The corpora in Zang's (2010) study are relatively bigger but are not quite comparable, as the TCCLC (Translated Chinese Children's Literature Corpus) and OCCLC (Original Chinese Children's Literature Corpus) do not follow same time span.

Yu (2014) and Jiang (2016) investigate the features of translated language in literature for children aged 3-6 years. Yu's (2014) study analyses vocabulary and grammar use in Chinese translated picture books compared to non-translated picture books, both in Chinese and English.

It ranges across variables like the frequency of content words and function words (including connectives and pronouns), long attributives, and passive constructions. The main findings of the study are that translations tend to use more frequent and concrete words; fewer idioms; more pronouns; more connectives; longer sentence lengths because of attributive *de*, and more passives. According to Yu (2014), simple words can ease reading difficulties for children, whereas more connectives, longer sentences and the unnatural use of passives in translation could burden children's reading comprehension. This is an important consideration for future translation, and ties in with the emphasis on target-audience acceptability in the translation of children's literature, discussed in Section 2.5.

Jiang (2016) investigates normalisation at lexical, syntactic and discourse levels. She finds the tendency of normalisation, but also deviation. Normalisation is most remarkable at lexical level and deviation is most detectable at syntactic and discourse levels. This study also discusses the disadvantages and advantages of normalisation and deviation. The biggest concern in relation to these two studies is the corpus type. Books for children aged 3-6 years are usually picture books, containing a large amount of illustrations and pictures, used to guide children to better understanding and to provide entertainment. The text and illustrations in picture books exist in a dialogic interaction as far as the meaning of the text is concerned (Lewis, 2001). Excluding visual information from the corpus may to some degree have influenced the findings of these two studies.

It is evident that there is considerable support for the hypothesis that translated language is significantly different from non-translated language in terms of explicitness, complexity and conventionality. However, despite the research on the features of translated language in Chinese translated from English, and some research on translated Chinese children's literature, comprehensive investigations of the features of translated language in translated Chinese children's literature are still very limited. This study addresses this research gap.

2.7 Conclusion

This chapter provided background to the concept of the features of translated language, the cognitive and socio-cognitive explanations offered for these features, text type and language pair as potential factors conditioning the realisation of these features, and the investigation of the features of translated language in children's literature in China. The aim of this discussion was to provide the theoretical background necessary to the study, and to provide a rationale for this study. In the next chapter, the research questions this study aims to answer will be

formulated against the background of the literature. The methodology used in this study will be discussed in more detail to provide information on the data collection and analysis used to answer these research questions.

Chapter 3: Methodology

3.1 Introduction

This chapter opens with an outline of the research questions arising from the literature review presented in Chapter 2 (Section 3.2). Against the background of these questions, the remainder of the chapter focuses on the corpus-linguistic method used to answer these questions. Section 3.3 focuses on the composition, compilation and processing of the comparable corpus of translated and non-translated Chinese children's literature used as dataset in this study. Subsequently, the operationalisations used as linguistic realisations of explicitation, simplification and normalisation are outlined and discussed in Section 3.4. The possibility that some of these linguistic realisations may also reflect transfer- or interference-related effects is also discussed. Section 3.5 outlines the methods used for data extraction, while Section 3.6 discusses the quantitative and qualitative analysis of the data.

3.2 Research questions

Against the background of the literature review in Chapter 2, this study aims to address the lack of research on the features of translated language in children's literature more generally, and in Chinese children's literature specifically. It aims to answer the following questions:

1. Does Chinese children's literature translated from English demonstrate evidence of the following three proposed recurrent features of translated language, in comparison with non-translated Chinese children's literature?
 - a. Increased explicitness
 - b. Simplification
 - c. Normalisation
2. What is the likely motivation for any differences in explicitness, complexity and conventionality observed between Chinese children's literature translated from English, in comparison with non-translated Chinese children's literature? In particular, is there evidence that source-text transfer, interference or "shining through" may account for observed differences?

In order to answer these questions in a systematic and quantifiable way, the study makes use of a corpus-linguistic methodology. The following sections explain in more detail what this methodology involves.

3.3 The Comparable Corpus of Translated and Non-translated Chinese Children's Literature: Corpus composition, text collection, and text processing

The well-known and widely used ZCTC (Zhejiang University Corpus of Translational Chinese) and LCMC (Lancaster Corpus of Mandarin Chinese), were compiled by Richard Xiao and his colleagues (McEnery & Xiao, 2004; Xiao & Hu, 2015) for contrastive and translational studies of English and Chinese. Another corpus that has been widely used is the CCTFC (Contemporary Chinese Translated Fiction Corpus), designed and constructed by Hu (2006) as a monolingual translational corpus of Chinese fictional texts. This corpus focuses on fiction (mainly for adults) translated from a number of source languages (eight at the time it was created) during the period 1980 to 2000 (Hu, 2007).

These corpora are commonly treated as representative of general contemporary Chinese writing and translation, including adult literature. They are, however, not suitable for the study of children's literature, since they do not include this text type. Partially because of this, the existing corpus-based studies of the features of translated Chinese children's literature discussed in Section 2.6 have used custom-built corpora compiled by the researchers themselves. These custom-built corpora suffer from some common drawbacks, in particular their limited size and thus poor representativeness. For this reason, it was deemed essential to compile a relatively large and comprehensive corpus of Chinese children's literature, both translated and non-translated, in order to answer the research questions outlined in Section 3.2. The following sections describe the composition, compilation and processing of the Translated Chinese Children's Literature Corpus (TCCLC) and Non-translated Chinese Children's Literature Corpus (NCCLC) used in this study.

3.3.1 Corpus composition

The TCCLC and NCCLC are constructed to be as comparable as possible, and in this study are used as subcorpora of a comparable corpus of translated and non-translated Chinese children's literature. The two subcorpora are constructed using Mandarin Chinese texts published in mainland China to ensure some degree of textual homogeneity (Xiao & Hu, 2015). The TCCLC contains 22 full texts of translated Chinese children's books, consisting of a total number of 1,168,137 tokens, while the NCCLC contains 20 full texts of children's books originally written in Chinese, with a total number of 1,215,259 tokens. The token count of each text varies

from 19,000 to 200,000, totalling 2.4 million tokens.⁵ The list of books included in each corpus is presented in detail in Appendix 1 and 2.

The decision to include full texts rather than text extracts was made to maintain the integrity of the data and ensure reliable frequency counts, as few linguistic features of a text are evenly distributed throughout a text (Saldanha & O'Brien, 2013, p. 74). Sampling may be particularly problematic where relatively less frequent linguistic features are investigated: "Frequency counts for common linguistic features are relatively stable across small samples (1,000 to 5,000 words) while frequency counts for rare features are less stable and require longer text samples to be reliably represented" (Biber, 1993, p. 249, as cited in Saldanha & O'Brien, 2013, p. 74).

In constructing the two subcorpora, several factors were kept in mind to ensure comparability. The books included are suggested as suitable reading for children aged from 7 to 11, either by the publishers/editors in the back-cover blurb of the book, or by booksellers in the classification by age groups in online bookstores. The decision to target books for slightly older children was partly motivated by the fact that pictures and illustrations are of vital importance in children's books for younger children, as visual material functions as a supplement to understanding the text, or even forms part of the content. A corpus of children's books for younger children therefore needs to be a multimodal corpus, including this visual material. Such corpora are more complex and time-consuming to construct than text-only corpora. However, in books for children aged above 7, the number of pictures and illustrations significantly decreases, and they can usually be removed without loss of meaning. Thus the two subcorpora created for this study include text characters only. The construction of a multimodal corpus is regarded as an important avenue for future research.

The design of the corpus avoids over-representing any individual author, translator or publisher. A contemporary timeframe of 2000-2017 was chosen for both subcorpora. However, this timeframe is slightly extended to 1998 for translated books, as a consequence of the fact that some famous translations by well-known translators do not, to my knowledge, have more recent editions. It should also be noted that for three of the translated books not much publishing information is available.

In surveying the available translated Chinese children's literature, it became evident that the most common translations are of classic fiction books for children. The composition of the

⁵ See Section 3.3.2 for a discussion of the approach followed to tokenise the Chinese texts.

TCCLC mirrors the reality of the translation industry in China, and reflects the dominance of children's book classics, for instance, *The Secret Garden* (Burnett, 1911), *Tuck Everlasting* (Babbitt, 1975) and *The Water Babies* (Kingsley, 1863). To match the composition of the TCCLC, the books included in the NCCLC are also considered classics of Chinese children's literature, for instance, 《我的妈妈是精灵》 *wǒ de māmā shì jīnglíng* (Chen, 2014, *My Mum is A Fairy*), 《魔法学校：小女巫》 *mófǎ xuéxiào xiǎo nǚwū* (Ge, 2015, *Magic School: A Little Witch*) and 《小老虎历险记》 *xiǎo lǎohǔ lìxiǎn jì* (Tang, 2013, *The Adventure of A Little Tiger*) (see Appendix 1 and 2 for a full list).⁶

3.3.2 Corpus compilation, processing and segmentation

Considering the time constraints associated with this project, the vast majority of texts included in the corpus were sourced from the Internet. Texts available on the Internet have the advantage of already having been digitised, thus limiting the need for time-consuming scanning and conversion of texts by Optical Character Recognition (OCR). However, even these digitised texts still require proofreading and manual correction to ensure that the electronic texts are accurate reflections of the original texts. A small number of texts were sourced by purchasing e-books from online bookstores in epub format. These were then converted into text files by the OCR module CamScanner (INTSIG, 2017). This transformation process resulted in a 1-3% error rate. These electronic text files were subsequently proofread and corrected manually in order to ensure accuracy. The corpora include running text only, and metadata are stored in a separate text file for easy retrieval.

Before the corpora could be used for analysis in the corpus-analysis software used in this study (WordSmith Tools 7.0, Scott, 2016), segmentation or tokenisation was necessary. Segmentation refers to “the process of segmenting text strings into word tokens, i.e. defining words (as opposed to characters) in a running text” (Xiao & Hu, 2015, p. 47). Segmentation is necessary for Chinese, because unlike English, which is written with spaces separating words, Chinese is presented as strings of characters without spaces. For segmentation, SegmentAnt 1.1.2 with the NLPIR/ICTCLAS engine (Anthony, 2017) was used, as it is free and accessible, with a high accuracy rate of 98.23%. The texts in the corpora were encoded in Unicode Transformation Format 8-Bit (UTF-8) for segmentation and then the segmented texts (see

⁶ The book titles are translated by the author.

Table 3.1: Features selected for investigation

Feature category	Linguistic indicator	Operationalisation	Examples of previous studies
Explication	Explicit signals of clausal relations	Conjunctions	Hu (2010), Hu & Zeng (2009), Huang (2007), Ke (2005), Mei (2015), Pápai (2004), Puurtinen (2004), Wang (2013), Xiao (2010), Xiao & Hu (2015), Yu (2014), Zang (2010)
	Increased explicitness of optional syntactic choices	Personal pronouns	Huang (2007, 2010), Jiménez-Crespo (2011), Mei (2015), Wang and Hu (2010), Xiao and Hu (2015), Zang (2010)
Simplification	Lexical variety	STTR (standardised type/token ratio)	Hu (2006), Mei (2015), Wang & Qin (2010), Xiao (2010), Xiao & Hu (2015)
	Syntactic complexity	Mean sentence length	Corpas Pastor (2008), Jiang (2016), Laviosa (1998), Xiao (2010)
Normalisation	Degree of normality	Modal particles	Jiang (2016), Mei (2015), Xiao & Hu (2015), Zang (2010)

These operationalisations are discussed in more detail in the following section, also drawing on the overviews presented in Chapter 2. A justification for the use of each operationalisation is provided, before more information about how the data collection was carried out is given.

3.5 Data collection

Data collection was conducted by using the Concord and Wordlist functions in the corpus analysis software WordSmith Tools 7.0 (Scott, 2016). The Wordlist function automatically generates word lists in both alphabetical and frequency order, allowing the researcher to identify high-frequency words in a corpus. The function of Concord is to retrieve all the occurrences of particular search terms or patterns in their immediate context and to display these in an easy-to-read format (Bowker, 2002).

3.5.1 Frequency of conjunctions

The system of conjunctions functions as “a complementary resource for creating and interpreting texts. It provides the resources for marking logico-semantic relationships that

obtain between text spans of varying extent, ranging from clauses within clause complexes to long spans of a paragraph or more” (Halliday & Matthiessen, 2004, pp. 538-539). The use of conjunctions is a way of setting up the logical relations between ideas, by virtue of the specific meanings of individual conjunctive items. As Halliday and Hasan (1976) describe it, “they are not primarily devices for reaching out into the preceding (or following) text, but they express certain meanings which presuppose the presence of other components in the discourse” (Halliday & Hasan, 1976, p. 226). Halliday and Matthiessen (2004) systematically divide conjunctions into three main types based on their semantic meanings and functions: elaboration, extension and enhancement.

Conjunctions can be treated as a realisation of explicitation because they can make the logico-semantic relations between propositions explicit in a number of different ways, for example, by restating some elements; clarifying ambiguity; adding extra information; and enhancing the meaning by qualifying time, place or manner.

In the case of Chinese, unlike its counterpart English, conjunctions are usually elliptic (Lü, 1998, p. 13). In example (1)⁷ the conjunction 若果 *rúguǒ* ‘if’, bracketed, can be left out before 你 *nǐ* ‘you’.

(1)

(rúguǒ) nǐ bù tīnghuà māma yào shēngqì le
(如果) 你 不 听话 妈妈 要 生气 了

If you don’t behave yourself, Mum will be angry.

Although the sentence without the conjunction is understandable and acceptable from the perspective of a native speaker, the inclusion of the conjunction would make the relationship between the two clauses more explicit. Various previous studies have used conjunctions as an operationalisation to investigate explicitation in translated Chinese, on the assumption that the more frequent use of conjunctions increases explicitness (see Hu, 2006; Xiao, 2010; Xiao & Hu, 2015; Wang, 2013).

⁷ All the examples provided in this study are formatted as follows: for examples of translated Chinese (from the TCCLC), the example in Chinese characters, a transcription in the Chinese phonetic alphabet, an English backtranslation, and the corresponding English source text are provided; for examples of non-translated Chinese, including texts from NCCLC, the example in Chinese characters, a transcription in the Chinese phonetic alphabet and English translations (by the author) are provided. The elements under discussion are underlined.

Chinese has a large variety of conjunctions (Xiao & Yue, 2009), which means that a thorough study of all conjunctions would be impossible in the limited scope of this study. Furthermore, it is known that conjunction use is strongly conditioned by register (Kruger & Van Rooy, 2012; Redelinghuys & Kruger, 2015), and since this study focuses on a specialised corpus of children's literature, the conjunctions in use might be particular to this genre. For this reason, a bottom-up method of identifying the most frequent conjunctions was selected. First, a word list was created for the corpus (including both subcorpora), using the Wordlist function in WordSmith Tools. From this, the ten most frequent tokens that could potentially be used as conjunctions were initially extracted. Table 3.2 shows these ten tokens in the combined corpus of translated and non-translated Chinese children's books.

Table 3.2: List of ten most frequent potentially conjunctive tokens in the combined corpus

No.	Chinese conjunction	Chinese phonetic alphabet	English translation ⁸	Frequency in corpus
1	和	<i>hé</i>	And	10,664
2	可	<i>kě</i>	but, yet	5,017
3	因为	<i>yīnwèi</i>	because, for, on account of	3,567
4	但	<i>dàn</i>	but, yet, still, nevertheless	3,031
5	可是	<i>kěshì</i>	but, yet, however	2,799
6	而	<i>ér</i>	and, but, for	2,646
7	却	<i>què</i>	but, yet, however, while	2,296
8	如果	<i>rúguǒ</i>	if, in case, in the event of, supposing that	1,924
9	不过	<i>bùguò</i>	but, nevertheless, however, only, except	1,895
10	然后	<i>ránhòu</i>	then, after that, afterwards	1,734

An important point is that this frequency criterion had to be based on the frequency of the item as clausal conjunctive, rather than the raw frequency of occurrence of the token. This was problematised by the fact that a number of the tokens have multiple uses. It was decided,

⁸ The English translations of the conjunctions in Table 3.2 and Table 3.3 are all taken from *The Chinese-English Dictionary* (Wu, 2010).

therefore, to narrow down the list in Table 3.2 to the five tokens most frequently used as clausal conjunctives. The selection process is described in more detail below.

The conjunction 和 *hé* ‘and’ is the most frequent conjunction in the combined corpus of translated and non-translated texts, with a frequency of 10,664 (normalised frequency: 4.47 per 1,000 words), more than double that of the second most frequent conjunction, 可 *kě* ‘but’ (frequency = 5,017, normalised frequency = 2.10 per 1,000 words). However, 和 *hé* ‘and’ typically functions to link words and phrases rather than clauses or sentences (Lü, 1999). It was therefore excluded from this study. The conjunction 可 *kě* ‘but’ has very diverse functions as it can act as a verb, conjunction or adverb. The frequency of 可 *kě* ‘but’ is inflated by the inclusion of its various functions as well as the mistaken inclusion of 可是 *kěshì* ‘but’, another listed conjunction. Manual analysis was therefore done to delete all non-conjunctive uses of 可 *kě* ‘but’. The concordances of 可是 *kěshì* ‘but’ were cleaned by deleting its adverbial use of stressing the tone, and adding the entries mistakenly included in the concordance of 可 *kě* ‘but’.

After the exclusion of 和 *hé* ‘and’, the sixth conjunction 而 *ér* ‘and/but’ became the fifth most frequent conjunction in the list. The high frequency of 而 *ér* ‘and/but’ is the result of the fact that it can be used both as a phrasal coordinator and a clausal coordinator. Manual sorting of the concordance entries was again conducted to remove phrasal conjunctive uses. After this, the frequency of 而 *ér* ‘and/but’ dropped dramatically from 2,646 to 1,577 (1.11 to 0.66 per 1000 words). Consequently, the seventh token 却 *què* ‘while’ in the word list was included to replace it. The concordances of 因为 *yīnwèi* ‘because’, 但 *dàn* ‘yet’ and 却 *què* ‘while’ were manually cleaned by deleting irrelevant cases. The final list of five conjunctions investigated is shown in Table 3.3. It is worthwhile noting that the conjunction 因为 *yīnwèi* ‘because’ is a causal conjunction while the rest four are all concessive conjunctions. This matter is taken up again in the discussion of the results in Section 4.2.1.

A full concordance of these five most frequent conjunctions was created for both the translated and non-translated subcorpora. With all irrelevant entries removed, these concordances were used to calculate the normalised frequency (per 1,000 words) of each conjunction, per file (text) in the corpus. The overall normalised frequency of all five conjunctions per file was also calculated. These normalised frequencies were used as the basis for statistical analysis.

Table 3.3: Selected conjunctions for investigation

No.	Chinese conjunction	Chinese phonetic alphabet	English translation	Frequency in corpus
1	因为	<i>yīnwèi</i>	because, for, on	3,508
2	可是	<i>kěshì</i>	but, yet, however	3,134
3	但	<i>dàn</i>	but, yet, still,	2,909
4	可	<i>kě</i>	but, yet	2,341
5	却	<i>què</i>	but, yet, however,	2,225

3.5.2 Frequency of personal pronouns

Personal pronouns, as a type of reference,⁹ refer to someone by specifying their function or role in the speech situation (Halliday & Hasan, 1976). The use of personal pronouns is a matter of identifiability by nature, as according to Halliday and Matthiessen (2004), pronouns occur where the given information can be retrieved from somewhere else by the receiver at the relevant point. In other words, personal pronouns indicate persons (or a person) in the common field of perception shared by speaker and receiver (Halliday & Matthiessen, 2004, pp. 550-551). Personal pronoun use increases the cohesiveness of a text, by anchoring reference in dependency relations in the discourse. In this indirect way, it may be seen as a linguistic manifestation of increased explicitness in translations if personal pronouns appear more frequently in translations than comparable non-translations.

More specifically related to explicitation is the fact that in Chinese some subject and object pronouns are often elliptic as long as the referential person of the pronoun is easy to identify from the context and will not cause ambiguity. In Lü's (1999) words: "Chinese does not use personal pronouns whenever it is not compulsory to be used; even though it might seem structurally incomplete, Chinese prefers no formalism" (Lü, 1999, p. 8).¹⁰ Xiao and Hu (2015) report one of the linguistic features of Chinese is "infrequent or non-compulsory use of

⁹ The term "reference" is consistent with Halliday and Hasan (1976) and Halliday and Matthiessen (2004).

¹⁰ Translation by the author.

referential components” (Xiao & Hu, 2015, p. 28). For example, one of the most commonly used greetings is shown in example (2).¹¹

(2)

nǐ chī le ma
(你) 吃 了 吗?
Have you eaten?

wǒ gāng chī le
(我) 刚 吃 了。
I have just eaten.

Here, the example omits the personal pronouns ‘you’ and ‘I’ at the positions of the sentence subject, which is idiomatic, and does not affect the success of communication. The inclusion of these pronouns would be acceptable, and would cause redundancy but also explicitness.

These kind of elliptical personal pronouns are broadly referred to as “optional pronouns” in this study. Optional personal pronouns clearly mark the subject or object of an action, and their more frequent use in translated Chinese, in comparison to non-translated Chinese, may demonstrate increased explicitness in the inclusion of optional elements. Studies by Wang and Hu (2010) and Xiao and Hu (2015) show that translated Chinese texts tend to include optional pronouns whereas native writers would be inclined to omit them.

However, leveraging this element of optionality to investigate increased explicitness in translation is complicated by the fact that pronouns are high frequency items, and there is no automated way of determining whether a particular instance is optional or obligatory. This would require time-consuming manual analysis. A compromise solution was therefore devised for this study: the frequency of personal pronouns in the two subcorpora was compared, and on the basis of differences identified, further qualitative analysis of concordances was done. All the Chinese personal pronouns in Table 3.4 were extracted using the Concord function in Wordsmith Tools. The frequencies of 俺们 *ǎnmen* ‘we’ and 您们 *nínmen* ‘you’ are 0 in both subcorpora, so they are not included in the analysis. These concordances were used to calculate the normalised frequency (per 1,000 words) of each personal pronoun, per file. The overall

¹¹ This greeting is typically used when two acquaintances meet each other around meal time. It is used in an informal conversation similar to the English greeting: “How are you?” “I’m good”.

normalised frequency of all personal pronouns per file was also calculated. These normalised frequencies were used as the basis for statistical analysis.

Table 3.4: Personal pronouns in Chinese

First person					
singular	我 <i>wǒ</i> ‘I’	俺 <i>ǎn</i> ‘I’ (colloquial)			
plural	我们 <i>wǒmen</i> ‘we’		咱 <i>zán</i> ‘we’ (colloquial)	咱们 <i>zánmen</i> ‘we’ (colloquial, inclusive)	俺们 <i>ǎnmen</i> ‘we’ (colloquial)
Second person					
singular	你 <i>nǐ</i> ‘you’		您 <i>nín</i> ‘you’ (honorific singular)		
plural	你们 <i>nǐmen</i> ‘you’		您们 <i>nínmen</i> ‘you’ (honorific plural)		
Third person					
singular	他 <i>tā</i> ‘he’	她 <i>tā</i> ‘she’	它 <i>tā</i> ‘it’		
plural	他们 <i>tāmen</i> ‘they’ (male)	她们 <i>tāmen</i> ‘they’ (female)	它们 <i>tāmen</i> ‘they’ (non-human)		

3.5.3 Standardised type/token ratio (STTR)

The measures of standardised type/token ratio (STTR) and mean sentence length were used to investigate lexical and syntactic simplification. Type-token ratio (TTR) refers to the ratio of the number of unique words (or types) and the number of running words (or tokens) (Kenny, 2001, p. 34). It reflects the variety of vocabulary used in a corpus: a lower TTR reflects a smaller range of vocabulary and more repetition. It is used as a “simple measure of the superficial lexical complexity of a text” (Munday, 1998, p. 4). In this sense, texts with a lower TTR have a smaller vocabulary range, and therefore might be simpler than texts with a higher TTR. However, TTR is often criticised for being sensitive to text length and therefore unreliable when analysing texts varying in length. Thus standardised TTR, with the default setting of standardisation for each 1,000 words was used as a measure of lexical variety. STTR is realised by calculating the ratio for consecutive 1,000-word chunks of texts and taking an average at the end (Kenny, 2001; Scott, 2016). The STTR for each text in the corpus was automatically calculated by Wordsmith Tools.

3.5.4 Mean sentence length

The use of mean sentence length as operationalisation departs from the intuitive idea that translators, especially translators of children's literature, are more likely to divide long sentences into several shorter sentences as a strategy to ease syntactic complexity. Several researchers, including Laviosa (1998), Xiao (2010) have used mean sentence length as an operationalisation to investigate simplification. However, as already noted in Section 2.2 and Section 2.3, shorter sentences do not necessarily correlate with decreased complexity. Shorter sentences may also be seen as more compact structures that are less explicit in meaning – and hence potentially more complex, rather than less. In other words, the correlation between shorter sentences and decreased syntactic complexity is not necessarily straightforward.

While being aware of these concerns, the present study treats mean sentence length as a potential measure of simplification in translation on the assumption that shorter sentences in children's literature are likely to be simpler sentences as well. While the definition of a sentence is somewhat contested (see Allerton, 1979; Nelson & Greenbaum, 2015), this study follows the practical definition of a sentence offered by Downing (2006, p. 272): "Grammatically, it is the highest unit and consists of one independent clause, or two or more related clauses. Orthographically and rhetorically, it is that unit which starts with a capital letter and ends with a full stop, question mark or exclamation mark." The calculation of mean sentence length (for each text) in Wordsmith Tools used in this study relies on an even simpler definition aligned with that of Pan's (1997): a written language segment that ends in a full stop, question mark or exclamation mark in written form.

3.5.5 Frequency of modal particles

Modal particles occur at the ends of groups of words which can be phrases, clauses or sentences (Chappell, 1991). They serve as an expression of the speaker's emotions or attitudes, including surprise, anxiety or suspicion. Only a few languages, including German, Japanese and Chinese make use of modal particles (Bross, 2012). Modal particles may be viewed as a unique feature of Chinese in contrast with English, which does not have a direct equivalent structure to Chinese modal particles. In English, the same functions can be realised by the use of auxiliaries, modal verbs, special word order or intonation. Modal particles in Chinese do not have meaning by themselves, but are context dependent, and thus each particle can be used in different contexts to express different emotions (Bross, 2012; Chappell, 1991).

According to Chappell (1991), modal particles are an essential component of colloquial language, particularly in informal contexts. In Chinese children’s literature, the use of modal particles is of particular interest for investigation because children’s literature tends to use modal particles more frequently than other types of writing to slow down reading speed and soften the tone (Zang, 2010). For instance, the expression 好啊! *hao a* ‘Sure!’ conveys the meaning of strong agreement while 好! *hao* ‘Good!’ without the modal particle expresses a much weaker emotion. The use of modal particles increases vividness in literature, which would be attractive to children and arouse resonance in them (Mei, 2015). The typical association of modal particles with Chinese children’s literature makes this operationalisation particularly suited to the investigation in this study.

Previous studies have investigated modal particles in relation to the features of translated language in Chinese (see Jiang, 2016; Mei, 2015; Xiao & Hu, 2015; Zang, 2010). There are two possibilities. If modal particles occur significantly more frequently in translated Chinese children’s books than in non-translated children’s books, this may be taken as evidence of normalisation or conventionalisation, with translators over-adjusting their translations to the typical norms of the target language. Alternatively, modal particles may occur significantly less frequently in the translated subcorpus than in the non-translated subcorpus. In this case, this may be ascribed to the effects of interference, specifically in the form of the Unique Items Hypothesis (see Section 2.2.4). As English lacks the counterpart of Chinese modal particles, the activation of English suppresses the activation of this “unique” feature of Chinese, leading to an under-representation of this feature in translated Chinese children’s books. A further possibility is that the lower frequency of these particles in translated Chinese might be ascribed to a different kind of conservatism on the part of translators: given their association with informal, colloquial language, translators may avoid them in favour of a more standardised, written style.

There are more than twenty types of modal particles in Chinese (Chao, 1968; Li & Thompson, 1981) and the focus of this study is on the most commonly used five: 吗 *ma*, 呢 *ne*, 吧 *ba*, 啊 *a*, 呀 *ya*. A bottom-up method of identifying the most commonly used modal particles was used. The most frequent modal particles were selected from the word list generated by the Wordlist function in WordSmith Tools of the combined corpus of translated and non-translated texts (see Table 3.5). It needs to be noted that the modal particle 了 *le* was excluded in this study. The token 了 *le* occurs with a frequency of 64,410 (27.02 per 1000 words). However,

了 *le* is not only used as a modal particle: it can be used after adverbs or adjectives as past tense marker to express that a certain action has finished. When it is used as a modal particle, it usually functions as a declarative marker. The high number of cases and the diversified use of 了 *le* made manual sorting of the concordance entries extremely time-consuming. Due to the limited scope of this study and time constraints, this modal particle was therefore excluded from this study. Further investigation of 了 *le* is foreseen as a future research possibility.

These particles share exactly the same forms with interjections in Chinese, but have different functions and positions. Interjections usually occur at the beginning of a sentence, but can be more flexible as well. They can form a sentence on their own. The concordances of these modal particles extracted from both corpora were manually cleaned by deleting irrelevant cases, where these tokens were used as interjections. These concordances were used to calculate the normalised frequency (per 1,000 words) of each modal particle, per file. The overall normalised frequency of all five modal particles combined per file was also calculated. These normalised frequencies were used as the basis for statistical analysis.

Table 3.5: Selected modal particles for investigation

No.	Modal particles	Chinese phonetic alphabet	Function ¹²	Frequency in corpus
1	呢	<i>ne</i>	Signals that a proposition is “contrary to expectations”; interrogative marker	4,582
2	吗	<i>ma</i>	Interrogative marker for polar “yes-no” questions	3,917
3	吧	<i>ba</i>	Codes suggestions; checks that hearer accepts the given proposition is a reasonable one	3,786
4	呀	<i>ya</i>	Prompts or urges hearers to do something	2,029
5	啊	<i>a</i>	Prompts or urges hearers to do something	1,755

As discussed above, modal particles do not have a lexical meaning of their own; instead their meaning is context-dependent. Applying the mood type classification (see Table 3.6) of Halliday and Matthiessen (2004) to Chinese, a configuration of modal particles with regard to

¹² The functions of modal particles are excerpted from Chappell (1991).

mood type can be produced (Huang & Liao, 2002; see Table 3.7). It should be noted that Table 3.7 does not attempt to summarise all the potential uses of each modal particle, but rather focuses on each modal particle's most prototypical usage.

Table 3.6: Mood type classification

mood type	indicative	declarative (statement)	declarative
			exclamative
		interrogative (question)	WH-interrogative (content questions)
			yes/no interrogative (polar questions)
			tag-questions
	imperative	(command)	

Table 3.7: Classification of modal particles by mood type in modern Chinese

declarative	呢 <i>ne</i> , 吧 <i>ba</i>
exclamative	啊 <i>a</i> , 呀 <i>ya</i>
interrogative	吗 <i>ma</i> , 呢 <i>ne</i> , 吧 <i>ba</i>
imperative	吧 <i>ba</i> , 啊 <i>a</i> , 呀 <i>ya</i>

In the presentation of the findings and discussion (see Section 4.2.4), the specific classification of the modal particles investigated in the study will be further refined building on this classification.

3.6 Data analysis

In order to answer the research questions, a comparison of the frequency/value of the selected operationalisations for the two subcorpora (the TCCLC and NCCLC) was needed. To evaluate whether the differences between the two groups were not just due to chance, a statistical test was required. The independent samples *t*-test is most commonly used; however, this test has three assumptions that need to be satisfied: independence of observations, normal distribution of data, and homogeneity of variance. The first assumption requires that each observation included in the analysis is independent of the others. In addition, the data should be

approximately normally distributed with the same variance (Baayen, 2008). The first condition was met in all cases, and to assess the second and third assumptions the following steps were followed for all operationalisations.

In the first step, a quantile-quantile (Q-Q) plot was produced and visually inspected to assess whether the data were sufficiently normally distributed to carry out the parametric *t*-test. In the second step, Levene's test was carried out to test the assumption of homogeneity of variance. If the test statistic (*p*-value) was larger than 0.05, then the equal variances assumption cannot be rejected.

In cases where the data were sufficiently normally distributed and the assumption of homogeneity of variance was met, subsequent descriptive statistics were presented using means as measure of central tendency, and standard deviation as measure of dispersion, and the *t*-test was used to determine whether the difference in means in the two subcorpora is statistically significant, with $p < 0.05$ set as the level of significance.

If these assumptions were not met, medians were used as measure of central tendency in reporting, and the interquartile range as measure of dispersion. The non-parametric two-samples Mann-Whitney *U*-test was used to assess the differences between the two subcorpora, with $p < 0.05$ regarded as a significant result. The Mann-Whitney *U*-test is a signed rank test suitable for non-normally distributed data (Baayen, 2008). Data in the two independent groups are combined and rank ordered together. There is no difference between the two groups when the values from the two groups are randomly mixed in the rank ordering, while there is a difference when they are clustered at opposite ends when combined (Corder & Foreman, 2011, p. 58).

In addition to the quantitative analysis, qualitative analysis of particular conjunctions, optional pronouns and modal particles was done in order to better understand the quantitative findings, and to further explore possible explanations for the observed findings.

Chapter 4: Findings and discussion

4.1 Introduction

This chapter reports and interprets the findings of the study. The central hypothesis of this study is that the translated subcorpus (the TCCLC) will demonstrate significantly increased explicitness, decreased complexity and increased conventionality in comparison to the non-translated subcorpus (the NCCLC). Section 4.2 presents a quantitative and qualitative analysis of each operationalisation selected for investigation in relation to explicitation, simplification and normalisation. The quantitative analysis aims to answer research question 1 (see Section 3.2), and to some degree research question 2. In order to develop a more fine-grained understanding of the findings and to further explore potential explanations (the aim of research question 2; see Section 3.2), further qualitative analysis was carried out. Potential evidence of transfer- or interference-related effects in the frequency and use of some of the operationalisations is also reported and discussed. Section 4.3 concludes this chapter by summarising key findings and conclusions.

4.2 Findings and discussion

4.2.1 Conjunctions

The quantile-quantile plot for all conjunctions (see Figure 1 in Appendix 3) indicates that the distribution of the data meets the assumption of normality (as discussed in Section 3.6). The results of Levene's test demonstrate that the assumption of homogeneity of variance is met ($F(1, 40) = 0.81, p = 0.37$). The results of the t -test confirm that there is a statistically significant difference between the TCCLC and the NCCLC ($t = 2.07, p < 0.05$), and the effect is in the direction expected: conjunctions overall occur more frequently in the translated than in the non-translated children's books. Figure 4.1 reflects that conjunctions are used at a mean normalised frequency of 6.27 per 1,000 words in the TCCLC, compared to 5.10 times per 1,000 words in the NCCLC.

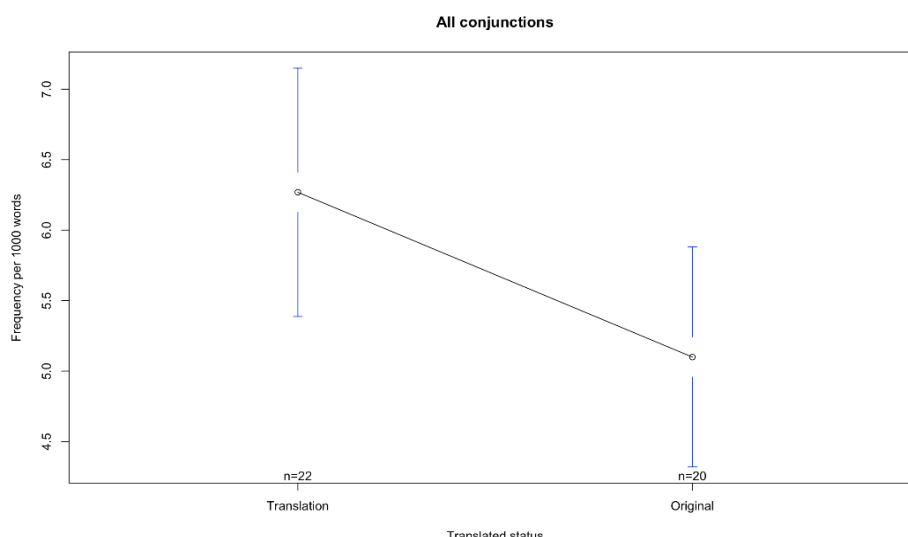


Figure 4.1: Normalised frequency of all conjunctions (per 1,000 words) by translated status

While the findings clearly support the hypothesis of increased explicitness in translation, in line with Xiao (2010) and others (see Section 2.3), an unanswered question is whether this increased explicitness is the consequence of translation-inherent explicitation, or whether there is potentially an interference or transfer effect. As discussed in Section 2.3, English is regarded as preferring a more explicit style than Chinese, and this source-language preference may be carried over in the translation (see also the discussion in Section 3.5.1). In an attempt to disambiguate these two possible causes for the increased frequency of conjunctions in the TCCLC, the five most frequent conjunctions were analysed individually (following the same procedure as for the overall analysis) to determine whether the tendency towards increased explicitness cuts across conjunction use more generally, or whether there are lexically specific effects that may potentially be ascribed to transfer.

The results of this analysis show there are no significant differences in the frequency of four out of the five most frequent conjunctions (see Table 4.1). These four conjunctions all represent a concessive logico-semantic relation.

Table 4.1: Conjunctions demonstrating no significant difference in frequency

Conjunction	Mean/median TCCLC	Mean/median NCCLC	Result of statistical test
但 <i>dàn</i> ‘yet’	Median 0.81	Median 0.74	$U = 259.50, p = 0.33$
可 <i>kě</i> ‘but’	Median 0.44	Median 0.60	$U = 196, p = 0.55$
可是 <i>kěshì</i> ‘but’	Median 0.90	Median 0.89	$U = 225.50, p = 0.90$
却 <i>què</i> ‘while’	Mean 0.88	Mean 0.84	$t = -0.15, p = 0.89$

This means that the significant overall difference in the frequency of conjunctions between the TCCLC and NCCLC is driven by a single conjunction, 因为 *yīnwèi* ‘because’, expressing causality, which demonstrates a highly significant difference in frequency between the two subcorpora. The boxplot in Figure 4.2 shows that in the TCCLC the conjunction 因为 *yīnwèi* ‘because’ occurs at a higher median frequency of 1.85 times per 1,000 words than in the NCCLC, where it occurs at a median frequency of 0.96 times per 1,000 words. Using the Mann-Whitney U -test, the difference between the two subcorpora is highly significant ($U = 366, p < 0.001$).

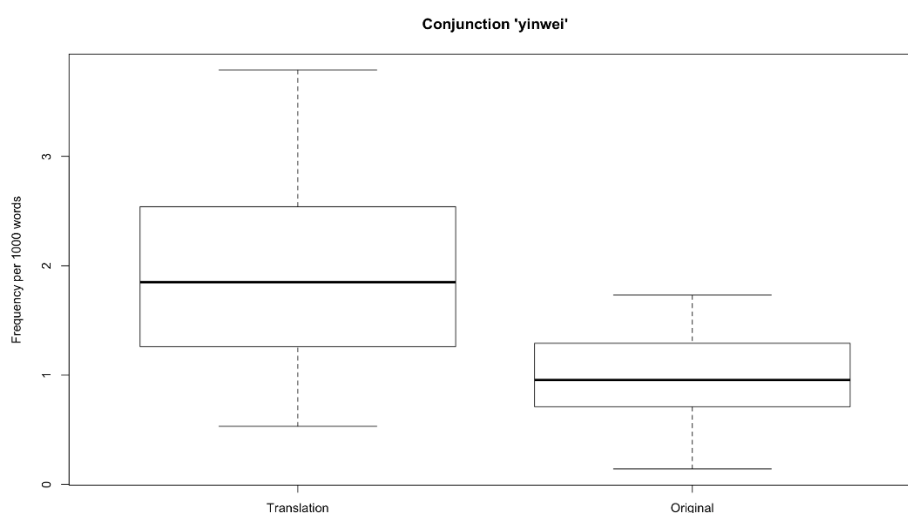


Figure 4.2: Normalised frequency of 因为 *yīnwèi* ‘because’ (per 1,000 words) by translated status

The fact that the tendency towards increased explicitness does not play out across all of the five most frequent conjunctions investigated, but is an effect associated with only one individual conjunction suggests that the increased explicitness is more likely related to a source-language transfer or interference effect, rather than an overall tendency towards increased explicitness. However, other interpretations may also be proposed. The following qualitative discussion considers three possible explanations for the increased frequency of 因为 *yīnwèi* ‘because’, in light of the functions, typical usage, and English translation equivalents of this conjunction.

The first proposed explanation centres on source-language transfer. English and Chinese are generally considered to demonstrate distinct stylistic preferences in terms of logico-semantic relations (Wang, 1984; Sun, 2013). Specifically, English prefers to use conjunctions to overtly link clauses, sentences or paragraphs, whereas Chinese prefers to use ordering to indicate the relationship between clauses, rather than overtly marking this relationship (Wang, 1984). This is particularly true in the case where the relationship in question involves causality, as Sun (2013) states conjunctions are unnecessary in two thirds of the cases where causality is involved in Chinese. The conjunction 因为 *yīnwèi* ‘because’ could be used both in forward linking and backward linking (see Li & Thompson, 1981 for an overview). By way of illustration, in Example (3a), 因为 *yīnwèi* ‘because’ is used to forward link the second clause in cause-and-effect order whereas in Example (4), 因为 *yīnwèi* ‘because’ is used to backward link the previous clause in an effect-and-cause order (also called cause preposition and cause postposition in Gao’s, 2013 terms). As put by Li and Thompson (1981, p. 641), “forward linking need not to be overtly marked at all, but can occur simply by virtue of the speaker’s intention”. Following this, the use of 因为 *yīnwèi* ‘because’ in Example (3a) could be deleted, as in Example (3b), as in (3b) the relationship between the two clauses could be inferred from the ordering of the context and do need to be signalled explicitly.

(3)

chén lǎoshī kěnéng shì yīnwèi zhàn lèi le biàn wānyāo zuòzài le ménbiān de yǐzi shàng

- (a) 陈 老师 可能 是 因为 站 累了, 便 弯腰 坐 在了 门边 的 椅子 上
Probably because Miss Chen was tired, she bent down and sat in the chair next door.

chén lǎoshī kěnéng shì zhàn lèi le biàn wānyāo zuòzài le ménbiān de yǐzi shàng

- (b) 陈 老师 可能 是 站 累了, 便 弯腰 坐 在了 门边 的 椅子 上
《魔法听诊器》 *mófǎtīngzhěnrì* ‘The Magic Stethoscopy’ (NCCLC)

(4)

Dàmàidì yǒu bùshǎo rénjiā de háizi bù dúshū yīnwèi méi qián

大麦地 有 不少 人家 的 孩子 不 读书, 因为 没 钱。

There are a lot of kids in Damaidi don't go to school, because they don't have money.

《青铜葵花》 *qīngtóngkuíhuā* 'QingtongKuihua' (NCCLC)

A closer analysis of the concordances of 因为 *yīnwèi* 'because' in the subcorpus of TCCLC reveals that the situation of forward linking offers a typical context in which translators opt to add 因为 *yīnwèi* 'because' even though it is not strictly required. Since causal conjunctions are optional in this context, it appears that translators add them because of the occurrence of English conjunctions in the source texts. The statistically higher frequency of 因为 *yīnwèi* 'because' in Chinese texts translated from English therefore might be motivated by the presence of its English equivalent. As this study is based on a comparable corpus consisting of translated and non-translated texts, this hypothesis cannot be tested using the current dataset, since the English source texts of the translations are not included in the analysis. However, a small-scale exploratory analysis comparing the translations with their English source texts was carried out in order to gain a better understanding of the findings. Example (5) and (6) are from the TCCLC, with the source texts recovered from online resources.

(5)

zài shùlín xiàmiàn yīnwèi zhīgàn géjué le yángguāng chàbùduō shì hēiyè le

在 树林 下面, 因为 枝干 隔绝 了 阳光, 差不多 是 黑夜 了 [TT]

Under the woods, because the branches blocked the sunshine, it was almost night there.

It was almost dark under the trees, for the branches shut out the daylight [ST]

《绿野仙踪》 *lǜyěxiānzōng* 'The Wizard of Oz' (TCCLC)

(6)

yěxǔ tōngwǎng nàge shí nián méi rén jiàn guò de huāyuán

也许 通往 那个 十 年 没 人 见 过 的 花园。 [TT]

Perhaps it led to the garden which had not been seen for ten years.

Perhaps it led into the garden which no one had seen for ten years. [ST]

yīnwèi mǎlì kě bùshì gè dǎnqiè de hái zǐ tā zǒudào lǜmén qián niǔdòng bǎshǒu

因为 玛丽 可 不是 个 胆怯 的 孩子,她 走到 绿门 前 扭动 把手。 [TT]

Since Mary was not a timid child, she went to the green door and turned the handle.

As she was not at all a timid child, Mary went to the green door and turned the handle. [ST]

《秘密花园》 mìmìhuāyuán 'The Secret Garden' (TCCLC)

These two examples are typical cause-and-effect sentences: the reason and the result are presented in sequence. There is an obvious logical and temporal order between the two parts. According to Li and Thompson (1981), in this situation, the conjunction 因为 yīnwèi 'because' is not necessary. Without it the logico-semantic relation between the clauses remains unambiguous. However, as evident in the examples, the translators include 因为 yīnwèi 'because' in their translations, most likely prompted by the fact that the English source texts include the causal conjunctions 'for' and 'as', which primes translators to include the Chinese conjunctions even though its use is not required or typical in this context. While this analysis is obviously limited and exploratory, the source-language transfer explanation is a possibility, and requires further quantitative investigation of translations in relation to their source texts, using a parallel corpus.

However, another potential explanation might be offered for the increased frequency of 因为 yīnwèi 'because' in the translation subcorpus, which is, in fact, in line with the notion of translation-inherent explicitation that is the result of constraints on the translator's language processing, imposed by the bilingual mode of language production (Kruger & Van Rooy, 2016; Laviosa, 2008). As for the most frequent concessive conjunctions (that do not demonstrate any significant differences in frequency in the TCCLC and the NCCLC), Chinese has a relatively large variety (also evident from Table 3.2) compared to English, which has a more limited number. In contrast, 因为 yīnwèi 'because' is the only highly frequent causal conjunction in Chinese (see Table 3.2 and Table 3.3). However, it also has a large variety of possible translation equivalents in English (see Zhao, 1995; Wang, 2007 for an overview). It can be translated as, amongst others, *because*, *since*, *as*, *for*, *on account of*, *as a result of*, and *as a consequence of*. It can be said that there is a many-to-one asymmetrical relation between the translation equivalents in English and Chinese. In addition, 因为 yīnwèi 'because' is the most frequent causal conjunction in the word list of the original subcorpus, which indicates that 因

为 *yīnwèi* ‘because’ is also the preferred causal conjunction in original Chinese children’s literature. Against this background, it appears reasonable to assume that when translators have to translate one of the many English conjunctions expressing causality, the most cognitively accessible Chinese equivalent is 因为 *yīnwèi* ‘because’. This conjunction is therefore selected for reasons of cognitive efficiency. At the same time, 因为 *yīnwèi* ‘because’ would also be the most conventional choice, and in this way, the relatively higher frequency of 因为 *yīnwèi* ‘because’ in the translation subcorpus may not only reflect increased explicitness, but also conventionality and potentially simplification, as 因为 *yīnwèi* ‘because’ is selected as an efficient and safe choice instead of other options.

A third plausible reason for the higher frequency of 因为 *yīnwèi* ‘because’ in the TCCLC is associated with the specific functions of children’s literature as a text type. According to Zhu and Hua (1992), children’s acquisition of compound sentences of causality in Chinese occurs later than the acquisition of other types of relations. Children start to use sentences expressing causality from age five to six, with limited conjunctions (Zhu & Hua, 1992) and it is only after the age of seven to eight that children consistently begin to understand the meaning of causality (Pi, 1980). The older children get, the more sensitive they become to the context cues indicating causal relations. Chen and Zhou (1989) investigated children’s understanding of the two types of causal sentences: forward and backward linking (see above discussion). The findings show that children have difficulty in understanding sentences with backward linking (Chen & Zhou, 1989, p. 44). Even if conjunctions are used to mark the causal relation, children’s understanding is still unsatisfactory (Chen & Zhou, 1989, p. 44). Against this background, it may be that translators choose to include causal conjunctions, and particularly the most frequent and most accessible causal conjunction 因为 *yīnwèi* ‘because’, to clarify the semantic relations for child readers. This interpretation would be in line with the hypothesis of increased explicitness in translation as a consequence of translators’ consciously making the text more accessible and explicit for readers.

4.2.2 Personal pronouns

The quantile-quantile plot (see Figure 2 in Appendix 3) confirms the normal distribution of the data, and the results of Levene’s test for equality of variance show that the assumption of equal variances is met ($F(1, 40) = 0.92, p = 0.34$). The result of the *t*-test for all pronouns shows there is a statistically highly significant difference in the overall frequency of pronouns between the

two subcorpora ($t = 3.65, p < 0.001$). The effect is in the expected direction: pronouns are more frequent in the translations than in the originals. It can be seen from Figure 4.3 that personal pronouns occur at a mean frequency of 83.89 times per 1,000 words in the TCCLC while the occurrence drops sharply to 62.47 times per 1,000 words in the NCCLC.

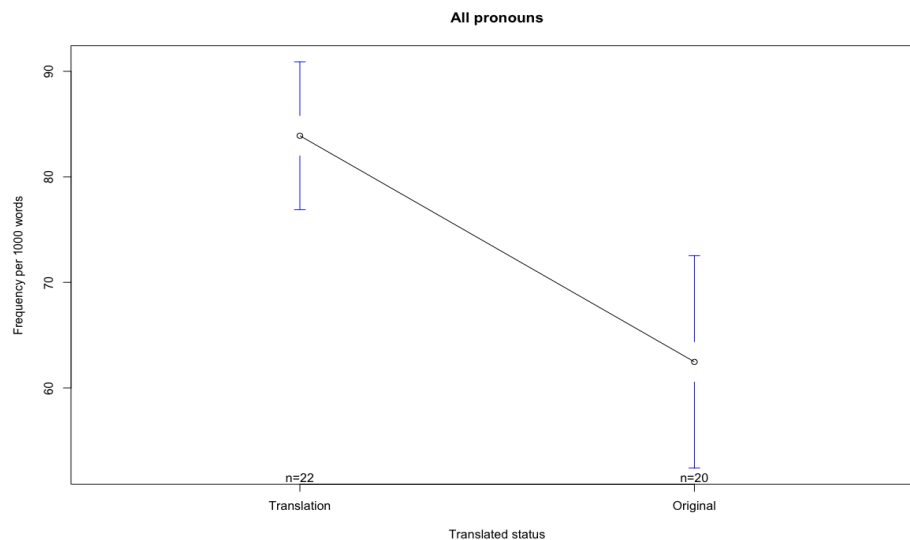


Figure 4.3: Normalised frequency of all pronouns (per 1,000 words) by translated status

As is the case for conjunctions (see Section 4.2.1), to answer the question of whether this finding is the consequence of translation-inherent explicitation, or whether there is potentially a source-language interference or transfer effect, each individual pronoun was analysed.

Table 4.2 summarises the findings for the individual pronouns demonstrating no significant difference in frequency, while Table 4.3 summarises the findings for pronouns that do demonstrate a significant difference in frequency. Following Table 4.3, a brief discussion of the significant differences in frequency for individual pronouns is presented.

Table 4.2: Pronouns demonstrating no significant difference in frequency

Pronoun	Mean/median TCCLC	Mean/median NCCLC	Result of statistical test
我 <i>wǒ</i> ‘I’	Median 22.07	Median 16.32	$U = 262, p = 0.30$
我们 <i>wǒmen</i> ‘we’	Median 4.18	Median 3.76	$U = 220, p = 1.00$
俺 <i>ǎn</i> ‘I’	Median 0.00	Median 0.00	$U = 196.50, p = 0.34$
咱 <i>zán</i> ‘we’	Median 0.00	Median 0.01	$U = 183.50, p = 0.32$
你 <i>nǐ</i> ‘you’	Mean 10.00	Mean 10.09	$t = -0.06, p = 0.95$
你们 <i>nǐmen</i> ‘you’	Mean 1.19	Mean 1.57	$t = -1.16, p = 0.25$
您 <i>nín</i> ‘you’	Median 0.05	Median 0.15	$U = 179.50, p = 0.31$

Table 4.3: Pronouns demonstrating a significant difference in frequency

Pronoun	Median TCCLC	Median NCCLC	Result of statistical test
他 <i>tā01</i> ¹³ ‘he’	15.52	10.54	$U = 294.50, p = 0.06$
他们 <i>tāmen01</i> ‘they’	5.87	2.50	$U = 320, p < 0.05$
她 <i>tā02</i> ‘she’	7.49	3.34	$U = 320, p < 0.05$
她们 <i>tāmen02</i> ‘they’	0.26	0.11	$U = 296, p = 0.06$
它 <i>tā03</i> ‘it’	3.82	1.55	$U = 336, p < 0.05$
它们 <i>tāmen03</i> ‘they’	1.03	0.32	$U = 361, p < 0.001$
咱们 <i>zánmen</i> ‘we’	0.00	0.10	$U = 119.50, p < 0.05$

As can be seen from Table 4.2 and Table 4.3, seven of the personal pronouns demonstrate significant differences in frequency. Furthermore, among the pronouns demonstrating significant differences, six are third-person pronouns for which the significant difference between the two subcorpora is in the hypothesised direction. In other words, except for the plural form of the first-person pronoun 咱们 *zánmen* ‘we’, all the other forms of first- and second-person pronouns do not show significant differences in frequency between the two

¹³ The Chinese phonetic alphabets of third person pronouns were numbered to distinguish them from each other in the statistical analysis environment in R, and the numbering system used in the analysis was retained here in the discussion of findings.

subcorpora whereas all the forms of third-person pronouns are significantly more frequent in the TCCLC compared to the NCCLC. The plural first-person pronoun 咱们 *zánmen* ‘we’ is the only pronoun where a significantly higher frequency is observed in the NCCLC rather than in the TCCLC, contrary to the hypothesis. Possible explanations for this are explored in the qualitative discussion below.

From the above, it is evident that there is a tendency of increased explicitness in the use of pronouns in the TCCLC, but this tendency plays out over a fairly restricted set of pronouns, namely, third-person pronouns. These findings echo those of Huang (2007), Wang and Hu (2010), and Xiao and Hu (2015). Similarly, Mei (2015) and Zang (2010) find that translational Chinese children’s literature demonstrates a higher overall frequency of personal pronouns than original Chinese.

Two (interrelated) explanations for this increased explicitness can be offered. The first explanation supports the notion of translation-inherent explicitation. Since pronoun use is, under certain conditions, optional in Chinese, translators may choose to add them to increase the explicitness of the text. This kind of explicitation is likely the result of translators’ awareness of their mediating communicative role in relation to readers and assumptions about their readership (Saldanha, 2008). The consideration of the target readers’ (and specifically child readers’) comprehension of reference may also explain the observed differences in the frequency patterns of first-, second- and third-person pronouns in the TCCLC compared to the NCCLC. Unlike the interactive first- and second-person pronouns, whose identity is recoverable from “the situational here and now” of the text, the identity of third-person pronouns must typically be recovered from within the text itself (Halliday & Matthiessen, 2004, pp. 551-552). In other words, the identity of first- and second-person pronouns are deictic. Their meaning is defined in the act of speaking: *I* is “the one speaking” and *you* is “the one(s) spoken to”. In contrast, third-person pronouns are typically used to refer to the person (or thing) in the preceding (sometimes following) text, and are thus anaphoric (Halliday & Matthiessen, 2004, p. 551). In this sense, the omission of third-person pronouns, especially beyond clause and sentence boundary, is more likely to cause ambiguity compared with first- and second-person pronouns. In the absence of the pronoun, readers (particularly child readers) may find it more effortful to retrieve the reference. Based on this consideration, even though the omission of pronouns within or beyond the sentence boundary is allowable, Chinese translators of children’s books may tend to add redundant third-person pronouns to more explicitly mark reference.

Translation-inherent explicitation would most prototypically be seen to occur in cases where there are no pronouns in the source text, but the translator adds it to explicitate the person involved as a retrospection or reminder, to facilitate readers' processing of the discourse. By way of illustration, in Example (7h), there is no source-text prompt or motivation for 他 *tā* 'him' since there is no corresponding 'him' in the source text. The translator adds 他 *tā* 'him' to specify to whom 'she asked wilder and wilder questions'. The same is also found in Example (7i), where the source text reads 'and hurried back to the flat below' while the translation reads 匆匆回到楼下他自己的套房 *cōngcōng huídào lóuxià tā zìjǐ de tàofáng* 'hurried back to his own suite downstairs'. Here, the translator adds 他自己 *tā zìjǐ* 'his¹⁴ own' in the translation. In both cases, it appears that there is an intention from the translator to transmit the information more explicitly by adding extra pronouns to avoid any potential for the readers to be confused.

(7)

ālún yífu bǎ tāngmǔ bào shàng lóu yímā zhèng děng zài nàlǐ
(a) 阿伦 姨夫 把 汤姆 抱 上 楼, 姨妈 正 等 在 那里。 [TT]

Uncle Alan carried Tom upstairs. Aunt was waiting there.

Uncle Alan carried Tom upstairs, to where his aunt was waiting. [ST]

ránhòu yífu yòu xiàlóu qù guānshàng huāyuán de mén ānwèi zhù zài dǐcéng de nàxiē
(b) 然后 姨夫 又 下楼 去 关上 花园 的 门, 安慰 住 在 底层 的 那些
fángkè

房客。 [TT]

Then uncle again go downstairs to close the garden door and reassure the tenants living ground-floor.

Then he went down again to shut the garden door and to reassure the ground-floor tenants.

[ST]

¹⁴ It needs to be noted that Chinese is a non-inflectional language, and thus does not have the classification of nominative and accusative case for pronouns. A typical possessive pronoun in Chinese is realised by the configuration of a pronoun followed by a suffix particle 的 *de*.

suǐhòu tā láidào zìjǐ zhù de èrlóu xiàng nàlǐ de qítā fángkè jiěshì shuō tā qīzǐ de
(c) 随后 他 来到 自己 住 的 二楼，向 那里 的 其他 房客 解释 说，他 妻子 的
wàisheng gāngcái mèngyóu láizhe

外甥 刚才 梦游 来着。 [TT]

Then he came to the second floor where he is living and explained to other tenants there that his wife's nephew had been sleep-walking just now.

Then he went up to his own floor and explained to the other tenant there that his wife's nephew had been sleep-walking. [ST]

zuìhòu tā shànglóu láidào bāsàiluòmiù tàitai de tàofáng
(d) 最后，他 上楼 来到 巴塞洛缪 太太 的 套房。 [TT]

Finally, he came upstairs to Mrs Bartholomew's suite.

Finally, he mounted to Mrs Bartholomew's flat. [ST]

tā fāxiàn tāde ménzhèng kāi zhe dàn shuān zhe tiěliàn
(e) 他 发现 她的 正门 开着，但 拴 着 铁链。 [TT]

He found that her front door was open but tied with a chain.

He found her with the front door open, but on a chain. [ST]

bāsàiluòmiù tàitai liǎnsè cāngbái húnshēn fādǒu bèi tā gāngcái tīngjiàn de hǎnjiàoshēng
(f) 巴塞洛缪 太太 脸色 苍白，浑身 发抖，被 她 刚才 听见 的 喊叫声
nòng de xīnfányiluàn

弄 得 心烦意乱。 [TT]

Mrs Bartholomew was pale and trembling. She was agitated by the crying she had just heard.

She was pale and trembling, and agitated by the crying she had heard. [ST]

tā tīng le tāde jiěshì dàn sìhū bìng bù xiāngxìn shènzhì sìhū bìng méiyǒu tīngdǒng
(g) 她 听 了他的 解释，但 似乎 并 不 相信，甚至 似乎 并 没有 听懂。 [TT]

She listened to his explanation, but she seemed not to believe, or even understand.

She listened to his explanation, but without seeming to believe or even to understand. [ST]

Tā wènle tā xǔduō yuèláiyuè mò míng qí miào de wèn tí ér qiě jiāng yī xiē wèn tí fǎn fù de
 (h) 她 问了他 许多 越来越 莫名其妙 的 问题, 而且 将 一些 问题 反复 地
 wèn le yòu wèn
 问 了 又 问。 [TT]

She asked him many increasingly weird questions and repeated some of them again and again.

She asked wilder and wilder questions, and asked the same ones again and again. [ST]

zuì hòu ā lún • jī tè sēn shī qù le nài xīn táng tū de xiàng tā dào le yī shēng wǎn ān
 (i) 最后, 阿伦 • 基特森 失去了 耐心, 唐突 地 向 她 道 了 一 声 晚安,
 cōng cōng huí dào lóu xià tā zì jǐ de tài fáng
 匆匆 回到 楼下 他 自己 的 套房。 [TT]

Finally, Alan Kitson lost patience, abruptly bade her a good night and hurried back to his own suite downstairs.

At last, Alan Kitson lost patience with her, bade her an abrupt good night and hurried back to the flat below. [ST]

《汤姆的午夜花园》 *tāng mǔ de wǎnyè huā yuán*, ‘Tom’s Midnight Garden’ (TCCLC)

A further possible interpretation relates to the notion of language-specific explication (Klaudy & Károly, 2005), a “shining through” or transfer-related increase in explicitness based on the stylistic differences identified between English and Chinese. Unlike English, which relies on a grammatical system of inflection, reference and conjunctions, Chinese has no inflection, and “infrequent and non-compulsory use of referential components, intra-sentential and inter-sentential conjunctions” (Xiao & Hu, 2015, p. 28). It can therefore be said that grammatically Chinese is an implicit language while English is an explicit language. English has been reported to demonstrate a higher frequency of personal pronouns compared to Chinese (Zhao & Shao, 2002). In other words, the more explicit style of English, where pronouns are not optional, may be transferred to Chinese translations.

In terms of reference in discourse, Chinese prefers “zero-pronouns” (Li & Thompson, 1981, p. 657). That is to say, in order to track a referent as a text unfolds, Chinese will either repeat the specified noun or leave the space where a pronoun should have been located blank, since referents that are clear from the discourse context do not need to be specified in Chinese (Dong, 2010; Li & Thompson, 1981; also see discussion in Section 3.5.2). Example (8) illustrates the former option. It is extracted from a narrative about a mother tiger named 力雅 *Liyǎ*, in which

the proper name 力雅 *Lìyǎ* occurs at every subject position instead of a third-person female pronoun 她 *tā* ‘she’. Example (9) illustrates the latter option, where the underlined blank space indicates there could have been a third-person pronoun 他 *tā* ‘he’ (and English would use a pronoun in each of these positions). Against this background, “language-specific explication” is therefore used here to refer to cases where there are pronouns in the source text, but these are not necessary in Chinese translation – yet the translator includes pronouns to explicitly specify the referent rather than leave it implicit, even though implicitness is allowed (and even preferred) in Chinese.

(8)

Lìyǎ shēnchū zhuǎzi pā de bǎ ròu dǎ diào yòu chòngzhe sìyǎngyuán liē kāi le zuǐ,

力雅 伸出 爪子，啪 地 把 肉 打 掉，又 冲着 饲养员 咧 开 了 嘴，
shuō Awu——awu——awu——wǒ yào shēngqì le

说：“啊呜——啊呜——啊呜——我 要 生 气 了！”

Liya stretched her claws, clapped down the meat and yelled at the feeder: “Awu—awu—awu—I’m going to be angry.”

Lìyǎ jìde Gūgu xiānshēng de xìn shàng shuō zhǐyào lǎohǔ yì shēngqì bùguǎn shì rén háishi

力雅 记得 咕咕 先生 的 信 上 说，只要 老虎一 生气，不管 是 人 还是
dòngwù, dōu huì xiàde fādǒu,

动物， 都 会 吓 得 发 抖。

Liya remembered that Mr. Gugu told her in the letter if a tiger gets angry, all the human beings and animals will shiver with fear.

yúshì Lìyǎ shēncháng bózi rang zìjǐ shēnshàng de máo dào shù le qǐ lái

于是 力雅 伸长 脖子，让 自己 身上 的 毛 倒 竖 了 起 来，

zhèyàng néng shǐ zìjǐ de gètóu kàn shàngqù dà xǔduō, fāqǐ píqì lái cái huì

这样 能 使 自己的 个 头 看 上 去 大 许 多，发 起 牌 气 来 才 会

xiǎnde gèng yǒu wēilì a wu wuwu—— Lìyǎ shēngqì de jiào

显 得 更 有 威 力 。“啊 呜 呜 呜 ——” 力雅 生 气 地 叫 。

Then *Liya* stretched her neck, letting the furs stand on end to make her look much bigger so that she appeared more powerful when getting angry. “A wu wuwu——” *Liya* howled with anger.

《小老虎历险记》‘*The Adventure of A Little Tiger*’ (NCCLC)

(9)

Dīlā yīnyīn gǎndào le nányán de kǒngjù tā juéde māma bùjiǔ yěyào lítā érqù
迪拉 隐隐 感到了 难言 的 恐惧，他 觉得 妈妈 不久 也要 离他 而去，
suǒyǐ pīnmìng zhēnxī hé māma gòngdù de zuìhòu shíguāng
所以 他 拼命 珍惜 和 妈妈 共度的 最后 时光。

Dila felt the unspeakable fear. He felt mum was going to be leaving him soon, so he cherished the last bit of moments with Mum.

Měitiān chúle bǔshí wài bù líkāi xuědòng bànbù yìzhí péi zài māma shēnpáng
他 每天 除了 捕食 外 不 离开 雪洞 半步， 一直 陪 在 妈妈 身旁。
He did not leave the cave except for the daily hunt, and stayed with Mum all the time.

《白狐迪拉与月亮石》‘White Fox Dila and Moon Stone’ (NCCLC)

Further pursuing the notion of “zero-pronouns” as part of the explanation for why third-person pronouns occur more frequently in the TCCLC than in the NCCLC, it should be noted that in comparison to first- and second-person pronouns, third-person pronouns may be more elliptical beyond clause and even sentence boundaries in narratives. For example, comparing Example (9) and (10), one notes that in the space of two sentences, the third-person pronoun 他 *tā* ‘he’ occurs only once after the introduction of the protagonist in Example (9) while the first-person pronoun 我 *wǒ* ‘I’ occurs four times in Example (10) (see the underlined). This is also a salient difference from English, in which third-person pronouns might be elliptical under strict conditions, but not beyond clause or sentence boundary. English makes use of long chains of reference after an antecedent (see Example (7a-e) for an illustration). In English-Chinese translation, translators might keep these chains of reference under the influence of the source text, even though it is appropriate in Chinese to omit third-person pronouns beyond the sentence boundary.

(10)

wǒ duì bàba de huà bù yǐwéirán xīnxiǎng wūyāzuǐ wǒ yòu búshì sānsuì de xiǎohái zěnmē
我 对 爸爸 的话 不以为然，心想： 乌鸦嘴， 我 又 不是 三岁 的 小孩，怎么
huì nàme bù xiǎoxīn ne
会 那么 不 小心 呢？

I don’t care about what Dad said, I am not a three-year-old child. How can I be that careless?

wǒ bù dāli bāba, xīnxiǎng: wúlùnrúhé wǒ dōu bùnéng hé hǎo yùnnqì zuòduì

我 不 搭 理 爸 爸 ， 心 想 ： 无 论 如 何 我 都 不 能 和 好 运 气 作 对 。

I don't reply to him. Anyway, I can't refuse good luck.

《口袋里的爸爸》 'Dad in the Pocket' (NCCLC)

However, determining to what degree this explanation accounts for the higher frequency of pronouns in the TCCLC is difficult in the current research design. Returning to the translation excerpt in Example (7), a closer analysis demonstrates that the translator does not apply a consistent strategy in dealing with third-person pronouns. In other words, there are traces of conventional zero-pronoun use, but also evidence of source-language related explicitation. For example, in (7b) and (7f), the original 'he' and 'she' at the beginning of sentences in the source text were translated as the proper names 姨夫 and 巴塞洛缪太太 rather than pronouns. In (7i) the literal translation of 'Alan Kitson lost patience with her' should be 阿伦·基特森(对她)失去了耐心 *ālún jītèsēn duì tā shīqù le nàixīn*. As can be seen in the Chinese translation, however, the translator omits 对她 *duì tā*, which is a typical case of pronoun omission in Chinese.

However, there is also evidence of explicitation. In the source text of (7a-e), Uncle Alan is first introduced into the discourse by means of proper name 'Uncle Alan'. After this introduction has been achieved, a chain of personal reference ('he') is used throughout. According to the "zero-pronouns" principle in Chinese, once the protagonist 阿伦 姨夫 *ālún yífū* has been established, the pronoun 他 *tā* in subject positions could be omitted, and typically would be, in non-translated Chinese writing. In other words, the subject 他 *tā* in (7c-e) is not obligatory because the agent performing the subsequent actions is perfectly clear from the discourse context and there is no need to specify the person with a pronoun. This is also the case in (7f), where the second 'she' in the source text is translated as 她 *tā*, which is also unnecessary since the subject 巴塞洛缪太太 *bāsàiluòmìu tài tai* has been established earlier within this sentence. This analysis suggests that it is the third-person pronouns in the source text that motivated the translator's choice to include them, as the occurrence of third-person pronouns in the translation corresponds with the occurrence of their counterparts in the source text.

To sum up, although the translator does make use of the conventional strategy to deal with reference, he does not do so consistently, most likely as a result of the interference of the source

language. Therefore, it might be argued that in this case, and many similar cases, source-language interference or transfer effects tend to trump normalisation effects, similar to the findings discussed in Section 2.3. As a consequence, the higher frequency of third-person pronouns in the English source text has been transferred to the translation, resulting in a higher frequency of these pronouns in the translations compared with the non-translations.

While there is therefore incontrovertible evidence for explicitation, translation-inherent and language-specific explicitation, or “shining through” combine in complex ways. Determining the exact role of the two types of explicitation is not possible within the scope of this study. Although the findings of this study are based on all personal pronouns (rather than only optional cases), it appears probable that at least part of the difference between the two subcorpora results from the fact that translators choose to use pronouns even where they are optional.

As pointed out above, a further finding requiring explanation is the findings for the pronoun 咱们 *zánmen* ‘we’. The boxplot in Figure 4.4 illustrates the single exception among the pronouns demonstrating a significant difference in frequency in the two subcorpora: the pronoun 咱们 *zánmen* ‘we’ (which is a low-frequency pronoun) occurs more frequently in the original subcorpus (at a median value of 0.10 per 1,000 words) than in the translation subcorpus, where the median frequency is 0.00.

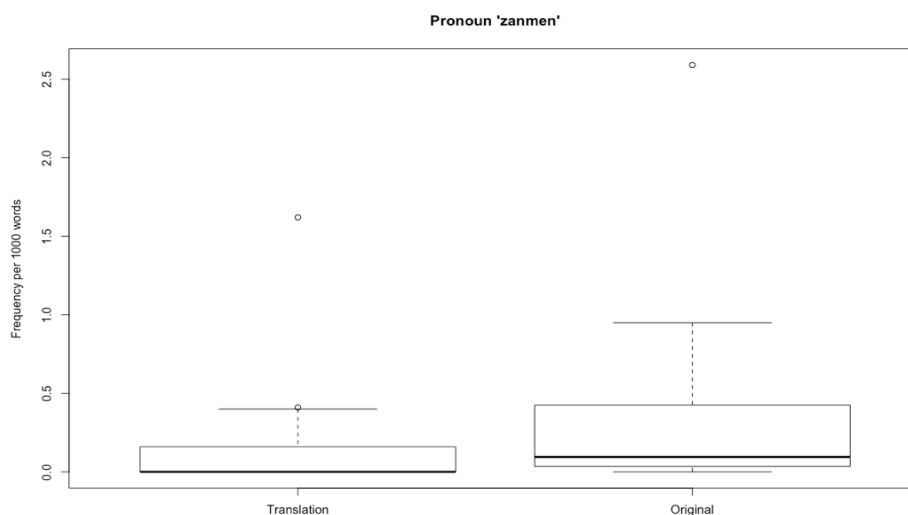


Figure 4.4: Normalised frequency of 咱们 *zánmen* ‘we’ (per 1,000 words) by translated status¹⁵

¹⁵The outliers in Figure 4.4, 4.6 and 4.8 are visualised to reflect the reality of the data (and were included in the analysis, which adopted appropriate statistical methods to deal with outlier values; see Section 3.6). These outliers

A possible explanation for this finding may be found in the Unique Items Hypothesis, which posits that items unique to the target language tend to be under-represented in the translation since there is no equivalent in the source text to prime its selection in the translation (see Section 2.2.4). 咱们 *zánmen* ‘we’ may be seen as such a unique item in Chinese, lacking a direct counterpart in English. It is a form of the plural first-person pronoun, similar to (the much higher-frequency) 我们 *wǒmen* ‘we’, for which no significant difference in frequency was found (see Table 4.2). However, 咱们 *zánmen* ‘we’ has a specialised function: it is typically used in colloquial spoken language as it originates in the North dialects in China (Chao, 1968; Lü, 1999), expressing an inclusive meaning, including the speaker and the receiver. In contrast, 我们 *wǒmen* ‘we’ is more widely used both in spoken and written language, expressing an inclusive or exclusive meaning, excluding the receiver (Chao, 1968; Li & Chen, 2009). In this sense, 我们 *wǒmen* ‘we’ is more likely to be seen as the equivalent of English ‘we’ since they share a similar function and semantic meaning. The use of 咱们 *zánmen* ‘we’ is likely to be more cautious due to its specialised meaning. By way of illustration, in Example (11) and (12), in the translation, 我们 *wǒmen* ‘we’ could be replaced by 咱们 *zánmen* ‘we’ since it is in conversation and the speaker intends to include both himself and the listener he is talking to. However, it can be seen that the original ‘us’ in the source text was translated as 我们 *wǒmen* ‘we’ rather than 咱们 *zánmen* ‘we’. Similarly, in 《夏洛的网》 *xiàluò de wǎng*, the translation of *Charlotte’s Web*, there are a number of cases where the use of 我们 *wǒmen* ‘we’ could be replaced by 咱们 *zánmen* ‘we’. However, the translator chooses to use 我们 *wǒmen* ‘we’ in the entire translation and there is no occurrence of 咱们 *zánmen* ‘we’. These examples confirm the likelihood of the Unique Items Hypothesis as possible explanation. The expression unique to the target language is underused in the translations.

(11)

hǎo le, wǒmen chī zǎofàn ba!

好了，我们吃早饭吧！

[TT]

All right, let’s have breakfast!

Let’s eat!

[ST]

《夏洛的网》 *xiàluò de wǎng* ‘*Charlotte’s Web*’ (TCCLC)

are also important in interpreting the findings, as is the case for Figure 4.8.

(12)

wǒ de tiān duōme chòu a wǒmen líkāi zhèer ba!

我 的 天 ！ 多 么 臭 啊 。 我们 离 开 这 儿 吧 ！ [TT]

Oh my god! How disgusting. Let's leave here.

Good night! What a stink! Let's get out of here! [ST]

《夏洛的网》*xiàluò de wǎng* 'Charlotte's Web' (TCCLC)

4.2.3 Standardised type-token ratio (STTR) and mean sentence length

The quantile-quantile plot (see Figure 3 in Appendix 3) for STTR shows a sufficiently normal distribution of the data. The results of Levene's test for the equality of variances demonstrate that the assumption of equal variances is met for the present analysis ($F(1, 40) = 2.40, p = 0.13$). Figure 4.5 shows that the mean STTR in translations and originals are 43.49 and 41.66, respectively. The results of the two-sample t -test, however, shows that the differences in means are not significant ($t = 1.31, p = 0.20$). However, it is observed that the mean value for STTR is higher for the translations than for the originals, contrary to the expectation that translations are relatively simpler than non-translations.

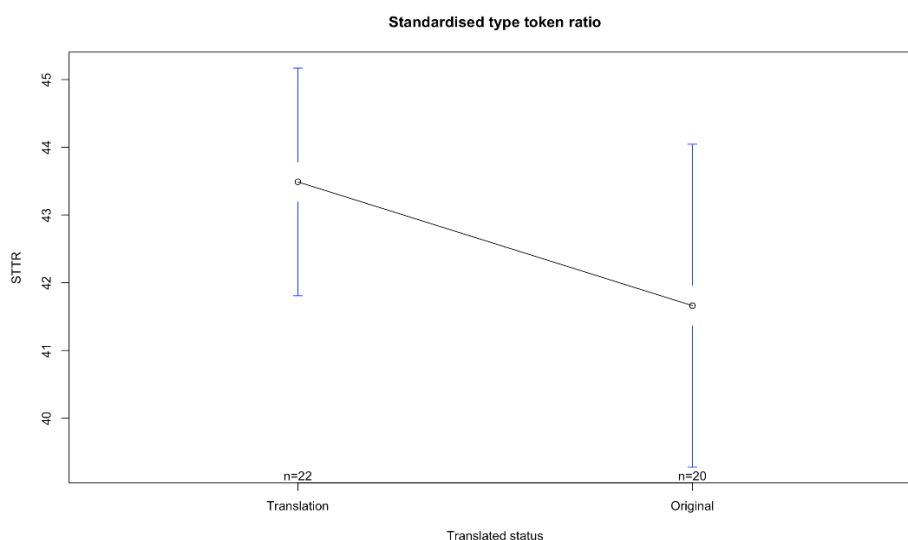


Figure 4.5: STTR by translated status

For mean sentence length, the quantile-quantile plot (see Figure 4 in Appendix 3) raises questions about the normal distribution of the data, predominantly because of the presence of outliers with very high values. Levene's test demonstrates that the assumption of equal

variances, however, is met for the present analysis ($F(1, 40) = 0.57, p = 0.45$). Figure 4.6 shows that there is hardly any difference between the medians of the translated and non-translated subcorpora, at 15.33 and 15.29 characters, respectively. Predictably, using the Mann-Whitney U -test, the difference between the two subcorpora is not significant ($U = 240, p = 0.63$).

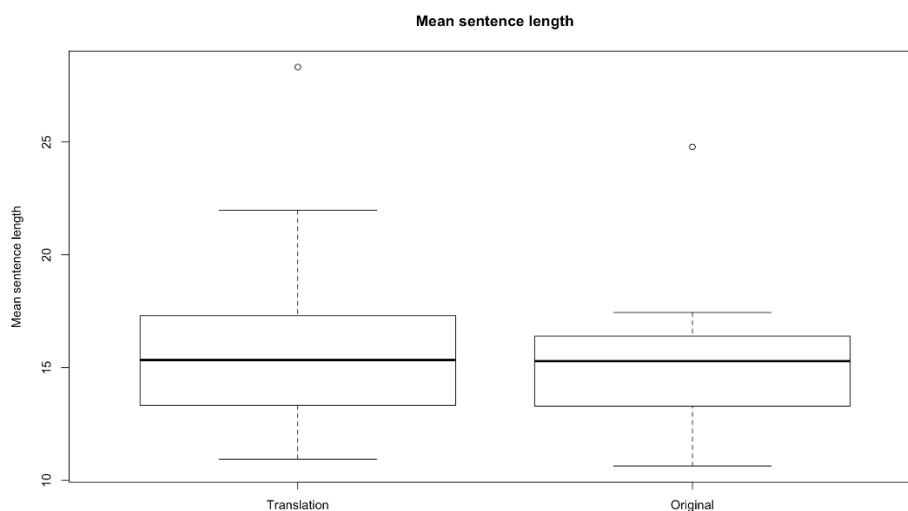


Figure 4.6: Mean sentence length (in characters) by translated status

STTR and mean sentence length were selected as operationalisations for the feature of simplification at the level of lexis and syntax, respectively. The findings for these two operationalisations confirm that there is no significant difference between the translations and non-translations in relation to either of these two features. Xiao (2010) and Xiao and Hu (2015) also find no significant difference in STTR between translated and non-translated corpora of general Chinese. The findings of the current study extend existing findings also to include children's books: as is the case in other text types, translated children's books in Chinese do not necessarily have a more limited vocabulary range than originals. Likewise, as far as sentence length is concerned, the translated and non-translated subcorpora have near identical median sentence lengths. This result fails to support Baker's (1996) assumption of simplification in translation, but is in line with the findings of Xiao (2010) and Xiao and Hu (2015) for Chinese. These findings also provide confirmation that translators of children's literature do not feel the pressure to simplify texts in translation more strongly than translators of other text types. Of course, it may well be that children's books are already relatively simple in terms of lexis and syntax, and there may therefore be a "floor" effect for simplification in translation: text types that already relatively simple in style are not simplified any further in

translation. Regardless of the explanation, the hypothesised feature of simplification does not seem to be an inherent feature of translations of children's literature in China.

4.2.4 Modal particles

The quantile-quantile plot (see Figure 5 in Appendix 3) indicates a sufficiently normal distribution of the data to carry out the parametric t -test, and the results of Levene's test show that the assumption of homogeneity of variance is met ($F(1, 40) = 0.04, p = 0.85$). Figure 4.7 shows that modal particles are more frequent overall in the originals (a mean of 7.74 per 1,000 words) than in the translations (6.25 per 1,000 words), in line with the findings of Xiao and Hu (2015). However, the t -test shows that this difference is not statistically significant ($t = -1.51, p = 0.14$).

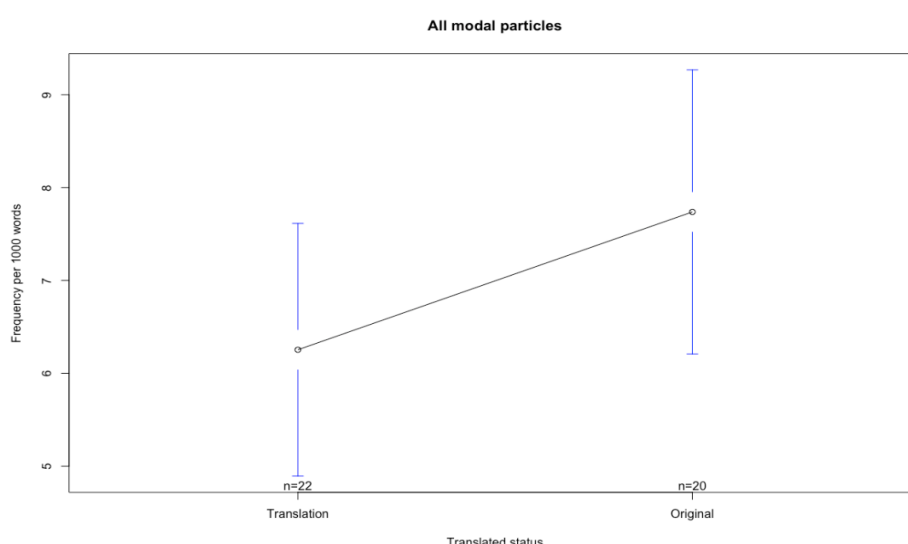


Figure 4.7: Normalised frequency of all modal particles (per 1,000 words) by translated status

While the overall findings therefore do not provide sufficient support to furnish a clear answer in support of either a conventionalisation or an interference effect, the latter appears to be more likely, based on the relative under-representation of modal particles in the TCCLC. Furthermore, it may also be possible that translators avoid modal particles because of their association with spoken language. A subsequent question is whether there are any differences for the individual modal particles that might shed light on potential explanations for the findings (using the same protocol as for the overall analysis).

The results show that three out of the five modal particles investigated do not demonstrate significant differences in the two subcorpora. Table 4.4 summarises the findings for the individual pronouns demonstrating no significant difference in frequency.

Table 4.4: Modal particles demonstrating no significant difference in frequency

Modal particle	Mean/median TCCLC	Mean/median NCCLC	Result of statistical test
吗 <i>ma</i>	Mean 1.99	Mean 1.55	$t = 1.33, p = 0.19$
吧 <i>ba</i>	Mean 1.49	Mean 1.80	$t = -1.13, p = 0.27$
啊 <i>a</i>	Median 0.58	Median 0.78	$U = 173.50, p = 0.25$

Two individual particles do demonstrate significant differences in the two subcorpora:

- the modal particle 呢 *ne* is significantly more frequent in originals than in translations ($U = 139, p < 0.05$), in line with the overall trend (see Figure 4.8 and following discussion)
- the modal particle 呀 *ya* is significantly more frequent in originals than in translations ($U = 122, p < 0.05$), in line with the overall trend (see Figure 4.9 and following discussion).

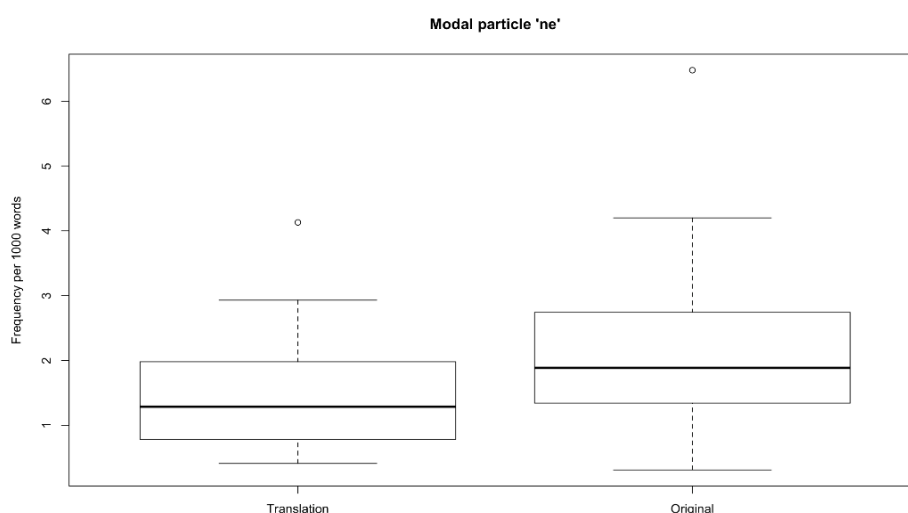


Figure 4.8: Normalised frequency of 呢 *ne* (per 1,000 words) by translated status

The boxplot in Figure 4.8 shows the median value for the (relatively frequent) modal particle 呢 *ne* in the TCCLC is 1.29 per 1,000 words, whereas the NCCLC has a higher median value of 1.89.

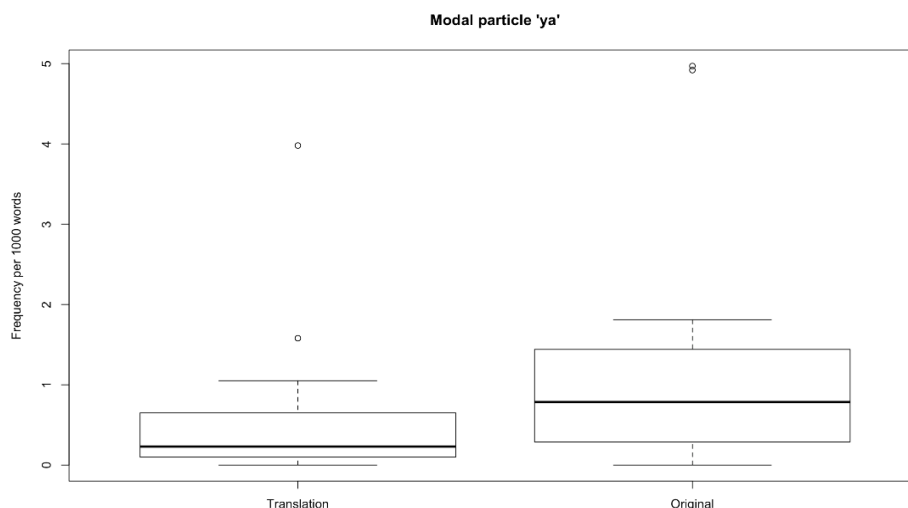


Figure 4.9: Normalised frequency of 呀 *ya* (per 1,000 words) by translated status

The boxplot in Figure 4.9 shows that the median values for the modal particle 呀 *ya* in the TCCLC and NCCLC are 0.23 and 0.79 per 1,000 words respectively. Clearly, the non-translated subcorpus has a higher median value. As discussed in Section 3.5.5, modal particles do not have a lexical meaning of their own; instead their meaning is context-dependent. In the following discussion, a more detailed discussion of the two modal particles demonstrating a significant difference in frequency is presented, based on an analysis of their association with sentence type or mood.

As shown in Table 3.7, 呢 *ne* may be used in both declaratives and interrogatives. When expressing declarative mood, it functions to signal that a proposition is contrary to what has been expected (Chappell, 1991) and shows the speaker's (or narrator's) wish to convince the interlocutor (Cao, 2005). For instance, in Example (13a), the speaker states that he definitely does not want to be a monk. 呢 *ne* is added to further stress the statement of 'don't want'. In this case, 呢 *ne* is optional, as seen in (13b).

(13)

dāngrán búshì wǒ cái bù xiǎng dāng héshàng ne

(a) 当然 不是，我 才 不想 当 和尚 呢。

Of course not, I don't want to be a monk.

dāngrán búshì wǒ cái bù xiǎng dāng héshàng

(b) 当然 不是，我 才 不想 当 和尚。

《口袋里的爸爸》*kǒudàilǐdebàba 'Dad in the Pocket'* (NCCLC)

In cases where 呢 *ne* is used to express interrogative mood, a distinction should be made between two different situations. In instances where there is a co-occurrence with a interrogative word (WH-elements in English), 呢 *ne* could be elliptical. This is because the interrogative mood is actually expressed by the interrogative word, not (or not completely) by the modal particle 呢 *ne* (Cao, 2005). 呢 *ne* functions to soften the tone so as to makes an enquiry readily acceptable. Therefore, the omission of 呢 *ne* would not cause significant loss of meaning and value. For instance, in Example (14a), the interrogative mood is basically conveyed by the interrogative word 哪儿 *nǎer* 'where', whereas the modal particle 呢 *ne* is merely used to soften a direct interrogation. As shown in Example (14b), without 呢 *ne* the sentence is still acceptable.

(14)

tā cáng dào nǎer qù le ne

(a) 她 藏 到 哪儿 去 了 呢？

Where has she hidden?

tā cáng dào nǎer qù le

(b) 她 藏 到 哪儿 去 了 ？

《青铜葵花》*qīngtóngkuīhuā 'Qingtong Kuihua'* in NCCLC

In other cases, where an interrogative word is omitted, the modal particle 呢 *ne* cannot be deleted. In some cases, an interrogative word can be omitted if, in the context, the omission will not cause any ambiguity. For instance, in Example (15a), from the answer given by the addressee, what 你家里人呢 *nǐ jiālǐrén ne* 'where are your families' means is actually 你家里人哪里去了 *nǐ jiālǐrén nǎli qù le* 'where have your families been' and here the interrogative word 哪里 *nǎli* 'where' has been omitted. In this case, the modal particle 呢 *ne* bears the

function of marking the interrogative tone, without which the interrogative mood disappears and the sentence is no longer correct (see Example (15b)). To sum up, except in interrogatives where an interrogative word is absent, 呢 *ne* may be elliptical (Wang, 2006).

(15)

- (a) nǐ jiālǐrén ne wǒ bàba māma jìn shān kǎnchái qù le
你 家里人 呢? 我 爸爸 妈妈 进 山 砍柴 去 了。

Where are your families?

- (b) nǐ jiālǐrén
*你 家里人? ¹⁶

《金猫历险记》 *jīnmāolìxiǎnjì* 'The Adventure of a Golden Cat' (NCCLC)

From the perspective of translation, since there is no direct equivalent in the English source text for the Chinese modal particle 呢 *ne*, especially in the instances where the use of a modal particle is not compulsory, translators' spontaneous cognitive response might be to translate without using it, since there is no trigger in the source text. In other words, on the one hand translators might not even notice that the choice to add 呢 *ne* exists. On the other hand, translators might be aware that they have the choice to include it or exclude it, but because of a lack of motivation in the source text, they might opt to not include 呢 *ne*.

This point can be illustrated by a comparison of two similar situations found in the NCCLC and TCCLC, where the non-translation (in Example (16)) does make use of 呢 *ne* while the translation (in Example (17) underlined) does not, even though adding 呢 *ne* sounds more idiomatic and typically would do so for a native writer. It appears likely that it is the lack of an equivalent in the English source texts that suppresses the use of 呢 *ne* in translation, and that source-text influence inhibits translators' use of the modal particle 呢 *ne*, resulting in a relatively lower frequency in the TCCLC.

¹⁶ The asterisk (*) indicates the sentence is regarded as ungrammatical.

(16)

qíshí wǒzìjǐ yě búshì zhème kàn de gǎnqíng yòu suàn shénme ne
其实我自己也不是这么看的，感情又算什么呢，
shuí dōu yǒu gǎnqíng kū qǐlái bíti bǐ yǎnlèi háiyào duō
谁都有感情，哭起来鼻涕比眼泪还要多。

Actually I don't think so either. How valuable are feelings? Everyone has feelings, and they cry with more snot than tears.

《我的妈妈是精灵》 *'My Mum is an Fairy'* (NCCLC)

(17)

tāngmǔ shàoye, wǒ nìngkěn búyào zhèfèner guāngróng shé huì bǎ wǒ de xiàba
汤姆少爷，我宁肯不要这份儿光荣。蛇会把我的下巴
gěi yǎo diào nà guāngróng hái suàn shénme
给咬掉，那光荣还算什么__？ [TT]

Master Tom, I would like to refuse this glory. The snake will bite my chin off and then how valuable is the glory?

Mars Tom, I doan' WANT no sich glory. Snake take 'n bite Jim's chin off, den WHAH is de glory? [ST]

《哈克贝利·芬历险记》 *hākèbèilì·fēnlìxiǎnjì 'Adventures of Huckleberry Finn'* (TCCLC)

The modal particles 呀 *ya* and 啊 *a* are generally interchangeable (Chappell, 1991). By using 呀 *ya* or 啊 *a* in exclamatives the speaker (or narrator) indicates that what he or she is experiencing is astonishing or surprising. As illustrated in Example (21a), by using the modal particle 呀 *ya* the speaker indicates that what she is seeing is unexpected for her, and this 呀 *ya* could be changed to 啊 *a*, as illustrated in Example (18b).

(18)

āiyā zhè bǎ xiǎo shànzi kě zhēn piàoliàng ya
(a) 哎呀，这把小扇子可真漂亮呀！

Oh, my! What a beautiful little fan!

āiyā zhè bǎ xiǎo shànzi kě zhēn piàoliàng a
(b) 哎呀，这把小扇子可真漂亮啊！

《没有风的扇子》 *méiyǒufēngdeshànzi 'A Fan without Wind'* (NCCLC)

When used in imperatives, 呀 *ya* has a hortatory use in prompting or urging the interlocutor to do something (Chappell, 1991). In Example (19a), the modal particle 呀 *ya* is used by the speaker to give a command to urge the listener to take action. Again, 呀 *ya* could be replaced by 啊 *a*, as in Example (19b).

(19)

- guòlái ya lǎohuān héshǔ hǎndào
(a) “过来 呀，老獾！” 河鼠 喊道。 [TT]
“Come here old Badger” the Rat shouted.

“Come on, old Badger!” shouted the Rat. [ST]

- guòlái a lǎohuān héshǔ hǎndào
(b) “过来 啊，老獾！” 河鼠 喊道。

《柳林风声》 *liǔlínfēngshēng* ‘*The Wind in the Willows*’ (TCCLC)

The use of 呀 *ya* and 啊 *a* in native, non-translated writing shows no significant difference in frequency, as they have nearly identical medians in the NCCLC (呀 *ya*: 0.79; 啊 *a*: 0.78). When compared with translations, however, a significant difference in frequency only occurs for 呀 *ya*, which is more frequent in the NCCLC than in the TCCLC.

The most likely explanation for this finding is that there may be some influence of writers’ and translators’ personal preferences. The quantile-quantile plot for 呀 *ya* shows non-normal distribution of the data (see Figure 6 in Appendix 3), with outliers in both the TCCLC and NCCLC. The most frequent use of 呀 *ya* in the TCCLC occurs in 红头发安妮 *hóng tóufà ānni* ‘*Anne of Green Gables*’, with a frequency of 3.98 times per 1,000 words, while the most frequent use of 呀 *ya* in the NCCLC occurs in the book 下次开船港 *xiàcì kāichuán gǎng* ‘*Next Time Depart Bay*’ (4.92 times per 1,000 words) and 大林小林 *dàlínxiǎolín* ‘*Big Lin and Little Lin*’ (4.97 times per 1,000 words). The frequency of 呀 *ya* in the rest of the texts in the two subcorpora vary from 0 to 1.81 times per 1,000 words. While the Unique Items Hypothesis, as discussed above, in all likelihood accounts for the significantly higher frequency of 呀 *ya* in the original subcorpus than in the translation subcorpus, there also appears to be some effect of individual preference by authors (and translators) in selecting 呀 *ya* rather than 啊 *a*.

In sum, the results for modal particles therefore fail to provide supporting evidence for the feature of normalisation. The relatively more frequent occurrence of modal particles in the originals than the translations suggests that the Unique Items Hypothesis, a form of “negative transfer” might have an effect. Since English does not have a structure that is directly equivalent in form and function to Chinese modal particles, modal particles can be regarded as a unique feature of the target language. The activation of the source language, English, suppresses the activation of this “unique” feature of Chinese, which has no counterpart in English. Consequently, modal particles tend to be under-represented in the translation subcorpus compared to the non-translated subcorpus.

4.3 Summary and conclusion

This chapter presented the findings of the study, and offered detailed discussion and interpretation of these findings. Two operationalisations were selected to investigate whether translated Chinese children’s literature demonstrates **increased explicitness** in comparison with non-translated Chinese children’s literature: conjunctions and optional personal pronouns. There is a statistically significant difference in the overall frequency of both conjunctions and pronouns, in the direction predicted: the translated subcorpus makes more frequent use of both these features. There is thus substantial support for the prediction that translated Chinese children’s literature tends towards increased explicitness of lexicogrammatical encoding in comparison to non-translated Chinese children’s literature. What is less clear is whether this difference can be ascribed to “translation-inherent” explicitation, or whether the increased explicitness is in fact a result of transfer or interference effects as a consequence of the fact that English prefers more explicit lexicogrammatical encodings than Chinese. In an attempt to tease apart these two explanations, individual conjunctions and personal pronouns were analysed in more detail.

As the findings in Section 4.2.1 demonstrate, the significant overall difference in the frequency of conjunctions between the two subcorpora is actually driven by the highly significant difference of only one conjunction: 因为 *yīnwèi* ‘because’, indicating causality. The fact that the tendency towards increased explicitness does not play out across all of the conjunctions investigated, but is an effect associated with only one individual conjunction suggests that the increased explicitness is more likely related to a source-language transfer or interference effect, rather than an overall tendency towards increased explicitness. Detailed further analysis demonstrates that at least three interwoven factors play a role. When 因为 *yīnwèi* ‘because’ is

used to forward link cause-and-effect clauses, the use of 因为 *yīnwèi* ‘because’ is optional, because the causal relationship can be inferred from the context. A close analysis of concordances shows that translators’ preference for including the optional 因为 *yīnwèi* ‘because’ does appear to form part of the reason for the increased frequency of this conjunction in the translated subcorpus: translators tend to include the conjunction 因为 *yīnwèi* ‘because’ to explicitate the relationship even if it is not necessary. However, there is also a clear transfer effect: the appearance of 因为 *yīnwèi* ‘because’ corresponds to the occurrence of its counterpart in English. In other words, the explicit style of marking clausal relationships in English has been transferred to the translation in Chinese. Lastly, translators’ intention to explicitly signal the relationship for child readers to promote their comprehension of causality may also form part of the explanation for these findings.

As far as the second operationalisation, personal pronouns, is concerned, the findings in Section 4.2.2 provide support for the explicitation hypothesis, in that personal pronouns demonstrate a statistically significant higher overall frequency in the TCCLC than in the NCCLC. Moreover, this higher frequency is driven by third-person pronouns, with all six third-person pronouns significantly (or almost significantly) more frequent in the translation subcorpus compared to the non-translation subcorpus. Two interrelated explanations in respect to explicitation may be offered for this tendency: translation-inherent and language-specific explicitation (which may be seen as a form of transfer or interference). On the one hand, translators try to make the third party involved explicit since unlike the first- and second-person pronouns, recoverable from the speech situation, the recovery of the referents of third-person pronouns is more likely textually based. The identification of third-person pronouns may require more cognitive effort from readers, and may be particularly challenging for child readers in the absence of an explicit pronoun. Therefore, translators feel the duty to add explicit pronouns for their target readers, even to the extent of adding extra pronouns. On the other hand, English has a more explicit style in using referential elements than Chinese, as English is reported to make more use of pronouns than Chinese. In converting from an explicit language to a relatively implicit language, the explicit style may have been carried over in cases where a pronoun is no longer strictly required in Chinese, but is nevertheless included by translators. In other words, pronouns in the source text trigger translators’ choice to include them. The exception of 咱们 *zánmen* ‘we’, showing a significantly higher frequency in the original subcorpus than the translation subcorpus, is explained by its status as a unique item in Chinese.

In sum, the two operationalisations provide straightforward evidence that translated Chinese children's books prefer more explicit lexicogrammatical encoding than non-translated Chinese children's books. However, the source of this increased explicitness remains unclear: in both cases, the possibility of translation-inherent explicitation exists, but in both cases there is also evidence that the increased explicitness may also partially be the consequence of source-language interference. The latter explanation seems more plausible in explaining the frequency of the conjunction 因为 *yīnwèi* 'because' and third-person pronouns. In these cases, source-language interference seems to win out in accounting for the higher frequency of these elements in translated Chinese children's literature.

Lexical and syntactic simplification was investigated by the operationalisations of STTR and mean sentence length. The findings in Section 4.2.3 show neither of these two operationalisations demonstrate a significant difference between the two subcorpora. By these measures, translated Chinese children's books therefore are no less complex than children's books originally written in Chinese. These findings might be ascribed to the simple style of children's books: Children's books are already fairly simplified at the lexical and syntactic level, and this "floor" effect may mean that further simplification in translation is unlikely.

For normalisation, the frequency of modal particles was used as an operationalisation. Following the hypothesis of increased conventionality in translation, it was predicted that translated children's books might demonstrate a more frequent use of modal particles, as translators feel the pressure to conform their work to conventionalised usage patterns in the target language. As modal particles are a particularly distinct feature of Chinese children's literature, it was hypothesised that translators might overuse this feature in an attempt to meet the norms for the target-language genre. However, an alternative possibility is that modal particles may be less frequent in translated texts compared to non-translated texts. In this case, a possible explanation may be found in the Unique Items Hypothesis of Tirkkonen-Condit (2002), which is a kind of negative transfer that occurs when a target-language item lacks a direct equivalent in the source language.

The findings in Section 4.2.4 show that the original subcorpus has an observably more frequent use of modal particles than the translation subcorpus. While the overall difference between the two subcorpora is not significant, two out of five modal particles demonstrate significant differences in frequency, with a higher frequency in the original Chinese children's books than in the children's books translated from Chinese. These findings run counter to the

normalisation hypothesis, but is in line with the Unique Items Hypothesis. A small-scale concordance analysis with reference to the source texts suggests that it is the lack of a counterpart of modal particles in English that suppresses translators' use of them, even though modal particles may make translations more idiomatic.

To conclude, the findings of this study show the tendency of increased explicitness might be a distinct feature of Chinese children's books translated from English. However, the further analysis suggests that this explicitness is more likely to be the consequence of source-language transfer rather than inherent explicitation. The important role of source-language transfer is also evident in the use of modal particles, where the potential for normalisation is outweighed by the constraining effects of the source text. Lastly, translated Chinese children's books do not show decreased complexity.

Chapter 5: Conclusion

5.1 Introduction

The objective of this study was twofold. Firstly, it aimed to investigate whether there are significant differences in a set of features indexing explicitness, complexity and conventionality between translated and non-translated Chinese children's books. Secondly, it aimed to explore possible explanations the differences identified between the translated and non-translated children's books, with the aim of contributing to a better understanding of not only the motivations for the proposed features of translated language, but also the relationship between text type (in this case children's literature) and the features of translated language.

The features of translated language investigated in this study were explicitation, simplification and normalisation. Each of these features was operationalised by means of selected linguistic operationalisations. Moreover, following more recent developments in corpus-based translation studies, particular attention was given to transfer, interference or "shining through" as a potential feature of translated language, interacting with the features of explicitation, simplification and normalisation.

In this chapter, the methodology used in this study is briefly reviewed, a summary of the findings is provided and some conclusions are drawn. Implications of the findings, the limitations of this study and suggestions for future research are also discussed.

5.2 Review of the methodology

To answer the research questions posed, a large comparable corpus consisting of translated and non-translated Chinese children's books was compiled. The books selected in each subcorpus are classics, suitable for children aged from 7-11 years. Each operationalisation used in the study was selected and justified based on previous studies, and their potential relevance to an investigation to a study of translated children's literature in particular. For the operationalisation of conjunctions and modal particles, due to the limited scope of study, a bottom-up way was used to extract the most frequent conjunctions and modal particles in the combined corpus. The concord function in WordSmith Tools 7.0 (Scott, 2016) was used to calculate the frequency of conjunctions, pronouns and modal particles, while STTR and mean sentence length were calculated automatically in WordSmith Tools 7.0 (Scott, 2016).

After the necessary pre-processing of the data was completed, quantitative analysis was used to determine whether there are significant differences in the frequency of conjunctions, pronouns and modal particles in the two subcorpora, and whether STTR and mean sentence length demonstrate significantly different values. This analysis primarily aimed to answer research question 1, and to some degree question 2. For a more nuanced understanding of the results of the quantitative analysis, and to answer research question 2, a qualitative analysis was carried out. Since the corpus involved is a comparable corpus, a comparison between source texts and translation translations was not possible. However, a small-scale analysis with reference to the source texts was carried out to explore transfer, interference or “shining through” as a possible explanation for the findings.

5.3 Summary of findings

Section 4.3 provides a detailed summary of the findings of the study, and this discussion is not repeated here. Instead, this section highlights the broader trends that emerge from the findings of this study. In relation to the first research question, the findings of this study provide substantial evidence that Chinese children’s literature translated from English is more explicit in lexicogrammatical encoding than comparable non-translated texts. Decreased complexity and increased conventionality, however, do not appear to be significantly associated with translated Chinese children’s books, compared to original books – at least not as measured by the operationalisations investigated in this study.

In relation to the second research question, the analysis shows that the increased explicitness identified as a robust finding may be ascribed to numerous factors. There is evidence of translation-inherent explicitation, which may be the result of either cognitive constraints on the translator’s language processing, imposed by the bilingual mode of language production, or translators’ awareness of their communicative role with consideration to the target readers. For instance, the highly significant use of conjunction 因为 *yīnwèi* ‘because’ in the TCCLC could be seen as a consequence of the asymmetrical translation equivalents in the two languages. Chinese has a limited number of conjunctive expressions of causality compared to English, among which 因为 *yīnwèi* ‘because’ is the most frequent and thus the most cognitively accessible choice. Moreover, the causal relationship might potentially challenge child readers’ comprehension, given the fact that understanding of causal relations develop comparably late. Translators may wish to help their readers in understanding this relationship by adding explicit signals in the form of conjunctions. Similarly, translators may opt to add extra pronouns to ease

cognitive processing for their young readers, even though pronouns are often unnecessary in Chinese.

Additionally, divergence in stylistic preferences between English and Chinese also clearly account for some of the differences observed between the translated and non-translated children's books. Specifically, it appears that the stylistic preferences of English for more explicit lexicogrammatical encoding is often transferred to Chinese translations, leading to the “shining through” of English stylistic preferences in Chinese translations. The more explicit style of overtly marking relations between clauses and sentences also contributes to the higher frequency of 因为 *yīnwèi* ‘because’ in the TCCLC than in the NCCLC. This is reflected in the cases where the use of 因为 *yīnwèi* ‘because’ is not compulsory, but occurs because of translators’ literal translation of source texts. In a similar vein, a proportion of the more frequent use of third-person pronouns can be traced to their counterparts in the source texts.

Furthermore, source-language interference also plays a strong role in the frequency of modal particles in translated Chinese. In line with the Unique Items Hypothesis, translators’ use of modal particles tends to be constrained by the absence of this feature in the English source texts, which inhibits the selection of the feature unique to the target language.

From the above, it appears that source-language interference, transfer or “shining through” provides part of the explanation for the findings of this study. The source language has an influential effect in shaping tendencies of explicitation and normalisation in translated language in children's literature translation in China: sometimes helping these features along (as in the case of explicitation), and sometimes inhibiting them (as in the case of normalisation).

5.4 Implications, limitations and future studies

This study presented a corpus-based analysis of the features of translated language as they are evidenced in children's literature in China. Using a monolingual comparable corpus, it investigated the proposed features of explicitation, simplification and normalisation, as well as source-language transfer or “shining through”. This investigation of the features of translated language in Chinese children's literature may conceptually extend our knowledge concerning the generalisability of claims about the features of translated language, since studies of this kind have long focused on European languages and adult literature. In particular, this study makes a contribution to the growing body of research that considers the relationship between register and the features of translated language. The findings further strengthen the argument

that the features of translated language are more likely probabilistic tendencies of translation conditioned by text type as much as by the languages involved. The study highlights that the concept of “features of translated language” needs to be more cautiously qualified and differentiated in terms of language pairs and text type.

Beyond advancing our theoretical understanding of the nature of translated language and the effects of the translation process, the findings of this study in particular and findings of corpus-based translation studies in general may also have important implications for translator training and practice. For instance, the findings could be used to establish a theoretical base which would be a valuable source of translation strategies for translators of children’s literature. Based on the findings of this study, for example, the use of modal particles to make translated children’s literature more vivid and attractive could be a specific area targeted in advising translators of Chinese children’s books.

However, the findings of this study should be interpreted against the background of the limitations of the study, and suggest that more extensive investigations are required. The time and space constraints of the MRes thesis of necessity limits both the scope and the depth of investigation. First, in terms of corpus design, the use of a comparable corpus places limitations on the analysis of interference, transfer and “shining through” effects, since translations cannot be systematically compared with their source texts. A combination of a comparable corpus approach with a parallel corpus approach would allow for the more definite disentanglement of particularly the role of source-language interference or transfer. Furthermore, as mentioned in Chapter 3, the translation corpus was constructed to reflect the reality of translated children’s books in China while the non-translation corpus was constructed to be as comparable as possible to the translation corpus. However, the number of books included was relatively small and the contents were limited to classics. This somewhat limits the generalisability of the findings to this text type as a whole.

Second, more operationalisations at different linguistic levels could be investigated and would produce more evidence contributing to the aim of generalisability. In particular, future research should attempt to correlate findings for linguistic operationalisations associated with the different features, in order to determine statistically whether there are co-occurrence trends between different operationalisations that may lead us to theoretical refinement in understanding, for example, the complex relationship between explicitation and simplification. Last but not least, the different explanations that have been proposed for the features of

translated language in this study as well as other corpus-based studies are commonly intuition- or hypothesis-based. The only way to test these hypotheses is by experimental research. Corpora can point us in the right directions in seeking these explanations, but can ultimately not prove causation. For example, the promising proposal of the gravitational pull of the category prototype and highest-level schema as explanation for simplification and normalisation (Halverson, 2003) can only be tested by investigating translators' cognitive processing during the translation process. Likewise, translators' mental construction of potential readers can only be reconstructed through process-based research. The findings of corpus-based studies offer a starting point for the design of experimental research that can help us disentangle the explanations that have been proposed for the features of translated language. The combination of corpus findings with experimental research utilising methods like keystroke-logging and eye-tracking emerges as a crucial future area of research.

References

- Allerton, D. J. (1979). *Essentials of grammatical theory: A consensus view of syntax and morphology*. London: Routledge & Kegan Paul.
- Anthony, L. (2017). SegmentAnt (Version 1.1.2) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/>
- Baayen, R. H. (2008). *Analysing linguistic data: A practical introduction to statistics using R*. Cambridge/New York: Cambridge University Press.
- Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 233-250). Amsterdam: John Benjamins.
- Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In H. Somers (Ed.), *Terminology, LSP and translation: Studies in language engineering in honour of Juan C. Sager* (pp. 175-186). Amsterdam: John Benjamins.
- Baker, M. (1999). The role of corpora in investigating the linguistic behaviour of professional translators. *International Journal of Corpus Linguistics*, 4(2), 281–298.
- Baker, M. (2007). Patterns of idiomaticity in translated vs. non-translated text. *Belgian Journal of Linguistics*, 21(1), 11–21.
- Becher, V. (2010). Abandoning the notion of "translation-inherent" explicitation: Against a dogma of translation studies. *Across Languages and Cultures*, 11(1), 1-28. doi:10.1556/Acr.11.2010.1.1
- Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.
- Blum-Kulka, S. (1986). Shifts of cohesion and coherence in translation. In J. House & S. Blum-Kulka (Eds.), *Interlingual and intercultural communication: Discourse and cognition in translation and second language acquisition studies* (pp. 17-35). Tübingen: Günter Narr.
- Blum-Kulka, S., & Levenson, E. A. (1984). Universals of lexical simplification. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 119-139). London/New York: Longman.
- Bowker, L. (2002). *Computer-aided translation technology: A practical introduction*. Ottawa: University of Ottawa Press.
- Bross, F. (2012). German modal particles and the common ground. *Helikon*, 2, 182-209.
- Cappelle, B. (2012). English is less rich in manner-of-motion verbs when translated from French. *Across Languages and Cultures*, 13(2), 173-195.
- Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley/Los Angeles: University of California Press.
- Chappell, H. (1991). Strategies for the assertion of obviousness and disagreement in mandarin: A semantic study of the modal particle *me*. *Australian Journal of Linguistics*, 11(1), 39-65. doi: 10.1080/07268609108599451

- Chesterman, A. (2004). Hypotheses about translation universals. In G. Hansen, K. Malmkjaer, & D. Gile (Eds.), *Claims, changes and challenges in translation studies: Selected contributions from the EST Congress, Copenhagen 2001* (pp. 1-13). Amsterdam: John Benjamins.
- Corder, G. W., & Foreman, D. I. (2009). *Nonparametric statistics for non-statisticians: A step-by-step approach*. John Wiley & Sons. doi: 10.1002/9781118165881
- Corpas Pastor, G., Mitkov, R., Afzal, N., & Pekar, V. (2008). Translation universals: do they exist? A corpus-based NLP study of convergence and simplification. Retrieved from https://www.researchgate.net/publication/254434062_Translation_universals_do_they_exist_A_corpus-based_NLP_study_of_convergence_and_simplification
- Cvrček, V., & Chlumská, L. (2015). Simplification in translated Czech: A new approach to type-token ratio. *Russian Linguistics*, 39(3), 309-325. doi:10.1007/s11185-015-9151-8
- Delaere, I., De Sutter, G., & Plevoets, K. (2012). Is translated language more standardized than non-translated language? Using profile-based correspondence analysis for measuring linguistic distances between language varieties. *Target*, 24(2), 203-224.
- Downing, A., & Locke, P. (2006). *English grammar* (2nd ed.). London/New York: Routledge.
- Eskola, S. (2004). Untypical frequencies in translated language: A corpus-based study on a literary corpus of translated and non-translated Finnish. In A. Mauranen & P. Kujamäki (Eds.), *Translation universals: Do they exist?* (pp. 83-99). Amsterdam/Philadelphia John Benjamins.
- Frawley, W. (1984). Prolegomenon to a theory of translation. In W. Frawley (Ed.), *Translation: Literary, linguistic, and philosophical perspectives* (pp. 159-175). London: Associated University Presses.
- Gellerstam, M. (1986). Translationese in Swedish novels translated from English. In L. Wollin & H. Lindquist (Eds.), *Translation studies in Scandinavia* (pp. 88-95). Lund: CWK Gleerup.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar*. (3rd ed.). London: Arnold.
- Halverson, S. L. (2003). The cognitive basis of translation universals. *Target*, 15(2), 197-241.
- Hansen-Schirra, S. (2011). Between normalization and shining-through: Specific properties of English-German translations and their influence on the target language. In S. Kranich, V. Becher, S. Höder, & J. House (Eds.), *Multilingual discourse production: Diachronic and synchronic perspectives* (pp. 133-162). Amsterdam: John Benjamins
- House, J. (2008). Beyond intervention: Universals in translation? *Trans-Kom*, 1(1), 6-19.
- Jantunen, J. H. (2001). Synonymity and lexical simplification in translations: A corpus-based approach. *Across Languages and Cultures*, 2(1), 97-112.
- Jiménez-Crespo, M. A. (2011). The future of general tendencies in translation: Explicitation in web localization. *Target*, 23(1), 3-25. doi:10.1075/target.23.1.01jim
- Kenny, D. (2001). *Lexis and creativity in translation: A corpus-based study*. Manchester: St Jerome.

- Kenny, D. (2004). Parallel corpora and translation studies: Old questions, new perspectives? Reporting “that” in Gepcolt: A case study. In G. Barnbrook, P. Danielsson, & M. Mahlberg (Eds.), *Meaningful texts: The Extraction of semantic information from monolingual and multilingual corpora* (pp. 154-165). London: Continuum.
- Klaudy, K. (2008). Explicitation. In: Baker, M. & Saldanha, G. (Eds) *Routledge encyclopedia of translation studies* (2nd ed.) (pp. 104-108). London: Routledge.
- Klaudy, K., & Karoly, K. (2005). Implication in Translation: Empirical evidence for operational asymmetry in translation. *Across Languages and Cultures*, 6(1), 13-28. doi:10.1556/Acr.6.2005.1.2
- Kruger, H., & Van Rooy, B. (2012). Register and the features of translated language. *Across Languages and Cultures*, 13(1), 33-65. doi:10.1556/Acr.13.2012.1.3
- Kruger, H., & Van Rooy, B. (2016). Constrained language: A multidimensional analysis of translated English and a non-native indigenised variety of English. *English World-Wide*, 37(1), 26-57.
- Krüger, R. (2014). From S-explicitation to T-explicitation? Tracing the development of the explicitation concept. *Across Languages and Cultures*, 15(2), 153-175. doi:10.1556/Acr.15.2014.2.1
- Lanstyák, I., & Heltai, P. (2012). Universals in language contact and translation. *Across Languages and Cultures*, 13(1), 99-121. doi:10.1556/Acr.13.2012.1.6
- Lapshinova-Koltunski, E. (2015). Variation in translation: Evidence from corpora. In C. Fantinuoli & F. Zanettin (Eds.), *New directions in corpus-based translation studies* (pp. 93-114). Berlin: Language Science Press.
- Laviosa, S. (1998). Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta*, 43(4), 557-570.
- Laviosa, S. (2002). *Corpus-based translation studies: Theory, findings, applications*. Amsterdam: Rodopi.
- Laviosa, S. (2008). Universals. In Baker, M. & Saldanha, G. (Eds), *Routledge encyclopedia of translation studies* (2nd ed.) (pp. 306-310). London: Routledge.
- Laviosa-Braithwaite, S. (1997). Investigating simplification in an English comparable corpus of newspaper articles. In K. Klaudy & J. Kohn (Eds.). *Transfere necesse est. Proceedings of the second international conference on current trends in studies of translation and interpreting* (pp. 531-540). Budapest: Scholastica.
- Lee, D. Y. W. (2001). Genres, registers, text types, domain, and styles: Clarifying the concepts and navigating a path through the BNC jungle. *Language Learning & Technology*, 5(3), 37-72.
- Lefer, M.-A. (2012). Word-formation in translated language: The impact of language-pair specific features and genre variation. *Across Languages and Cultures*, 13(2), 145-172. doi:10.1556/Acr.13.2012.2.2
- Lefer, M.-A. & Vogeleer, S. (2013). Interference and normalization in genre-controlled multilingual corpora: Introduction. *Belgian Journal of Linguistics* 27: 1-21.

- Lewis, D. (2001). *Reading contemporary picturebooks: Picturing text*. London: Routledge.
- Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.
- Malmkjaer, K. (1997). Punctuation in Hans Christian Andersen's stories and in their translations into English. In F. Poyatos (Ed.), *Nonverbal communication and translation: New perspectives and challenges in literature, interpretation and the media* (pp. 151-162). Amsterdam: John Benjamins.
- Marco, J. (2009). Normalisation and the translation of phraseology in the COVALT corpus. *Meta*, 54(4), 842-856.
- Mauranen, A. (2000). Strange strings in translated language: A study on corpora. In M. Olohan (Ed.), *Intercultural faultlines: Research models in translation studies 1: Textual and cognitive aspects* (pp. 119 -141). Manchester: St. Jerome Publishing.
- Mauranen, A. (2004). Corpora, universals and interference. In A. Mauranen & P. Kujamäki (Eds.), *Translation universals: Do they exist?* (pp. 143-164). Amsterdam/ Philadelphia: John Benjamins.
- Mauranen, A., & Kujamäki, P. (Eds.). (2004). *Translation universals: Do they exist?* Amsterdam/Philadelphia: John Benjamins.
- McEnery, A., & Xiao, R. (2004). The Lancaster corpus of Mandarin Chinese: A corpus for monolingual and contrastive language study. In M. Lino, M. Xavier, F. Ferreire, R. Costa, & R. Silva (Eds.), *Proceedings of the fourth international conference on language resources and evaluation (LREC) 2004* (pp. 1175–1178). Lisbon: Centro Cultural de Belem.
- Munday, J. (1998). A computer-assisted approach to the analysis of translation shifts. *Meta*, 43(4), 542-556. doi:10.7202/003680ar
- Nelson, G., & Greenbaum, S. (2015). *An Introduction to English Grammar* (4th ed.). London/New York: Routledge.
- Neumann, S. (2014). Cross-linguistic register studies: Theoretical and methodological considerations. *Languages in Contrast*, 14(1), 35 –57
- O'Connell, E. (1999). Translating for children. In G. Anderman, & M. Rogers (Ed.), *Word, text, translation* (pp. 208-216). Clevedon: Multilingual Matters.
- Olohan, M. (2004). *Introducing corpora in translation studies*. London/New York: Routledge.
- Olohan, M., & Baker, M. (2000). Reporting that in translated English: Evidence for subconscious processes of explicitation? *Across Languages and Cultures*, 1(2), 141-158.
- Øverås, L. (1998). In search of the third code: An investigation of norms in literary translation. *Meta*, 43(4), 557-570.
- Pápai, V. (2004). Explicitation: A universal of translated text? In A. Mauranen & P. Kujamäki (Eds.), *Translation universals: Do they exist?* (pp. 143-164). Amsterdam/Philadelphia: John Benjamins.

- Puurtilinen, T. (1995). *Linguistic acceptability in translated children's literature*. Joensuu: University of Joensuu.
- Puurtilinen, T. (2003). Genre-specific features of translationese? Linguistic differences between translated and non-translated Finnish children's literature. *Literary & Linguistic Computing*, 18(4), 389-406.
- Puurtilinen, T. (2004). Explicitation of clausal relations: A corpus-based analysis of clause connectives in translated and non-translated Finnish children's literature. In A. Mauranen & P. Kujamäki (Eds.), *Translation universals: Do they exist?* (pp. 165-176). Amsterdam/Philadelphia: John Benjamins.
- Puurtilinen, T. (2006). Translating children's literature: Theoretical approaches and empirical studies. In G. Lathey (Ed.), *The translation of children's literature: A reader* (pp. 54-64). Clevedon: Multilingual Matters. (Reprinted from *Translation and the (Re) production of Culture*, pp. 273-283, by C. Robyns, Ed., 1994, Leuven: The CERA Chair for Translation, Communication and Cultures)
- Pym, A. (2005). Explaining explicitation. In K. Karoly & A. Foris (Eds.), *New trends in translation studies* (pp. 29-43). Budapest: Akademiai Kiado.
- Redelinghuys, K., & Kruger, H. (2015). Using the features of translated language to investigate translation expertise: A corpus-based study. *International Journal of Corpus Linguistics*, 20(3), 293-325.
- Rohdenburg, G. 1996. Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics* 7, 149-182. doi: 10.1515/cogl.1996.7.2.149
- Saldanha, G. (2008). Explicitation revisited- bringing the reader into the picture. *Trans-Kom*, 1(1), 20-35.
- Saldanha, G., & O'Brien, S. (2013). *Research methodologies in translation studies*. Manchester: St. Jerome Publishing
- Shavit, Z. (2006). Translation of children's literature In G. Lathey (Ed.), *The translation of children's literature: A reader* (pp. 25-53). Clevedon: Multilingual Matters. (Reprinted from *Poetics of children's literature*, pp.111-130, 1986, Athens: University of Georgia Press)
- Scott, M. (2016), WordSmith Tools (version 7.0) [Computer Software]. Stroud: Lexical Analysis Software. Available from <http://www.lexically.net/wordsmith/>
- Swan, M. (1980). *Practical English usage*. Oxford: Oxford University Press.
- Teich, E. (2003). *Cross-linguistic variation in system and text: A methodology for the investigation of translations and comparable texts*. Berlin: Mouton de Gruyter.
- Tirkkonen-Condit, S. (2002). Translationese - A myth or an empirical fact? A study into the linguistic identifiability of translated language. *Target*, 14(2), 207-220. doi:10.1075/target.14.2.02tir
- Tirkkonen-Condit, S. (2004). Unique items — over- or under-represented in translated language? In A. Mauranen & P. Kujamäki (Eds.), *Translation universals: Do they exist?* (pp. 177-184). Amsterdam/Philadelphia: John Benjamins.

- Toury, G. (1995). *Descriptive translation studies – and beyond*. Amsterdam: John Benjamins.
- Toury, G. (2004). Probabilistic explanations in translation studies: Welcome as they are, would they qualify as universals? In A. Mauranen & P. Kujamäki (Eds.), *Translation universals: Do they exist?* (pp. 15-32) Amsterdam/Philadelphia: John Benjamins.
- Toury, G. (2012). *Descriptive translation studies – and beyond* (2nd ed.). Amsterdam John Benjamins.
- Ulrych, M., & Murphy, A. (2008). Descriptive translation studies and the use of corpora: Investigating mediation universals. In C. T. Torsello, K. Ackerley, & E. Castello (Eds.), *Corpora for university language teachers*. Bern: Peter Lang.
- Van Coillie, J., & Verschueren, W. P. (Eds.). (2014). *Children's literature in translation: Challenges and strategies*. London: Routledge.
- Vanderauwera, R. (1985). *Dutch novels translated into English: The transformation of a "minority" literature*. Amsterdam: Rodopi.
- Wang, K., & Qin, H. (2010). A parallel corpus-based study of translational Chinese. In R. Xiao (Ed.), *Using corpora in contrastive and translation studies* (pp. 164–181). Newcastle: Cambridge Scholars Publishing.
- Xiao, R. (2010). How different is translated Chinese from native Chinese? A corpus-based study of translation universals. *International Journal of Corpus Linguistics*, 15(1), 5-35.
- Xiao, R. (2011). Word clusters and reformulation markers in Chinese and English. *Languages in Contrast*, 11(2), 145-171. doi:10.1075/lic.11.2.01xia
- Xiao, R., & Hu, X. (2015). *Corpus-based studies of translational Chinese in English-Chinese translation*. In D. Li (Series Ed.), *New frontiers in translation studies*. doi: 10.1007/978-3-642-41363-6.
- Xiao, R., McEnery, T., & Qian, Y. (2006). Passive constructions in English and Chinese: A corpus-based contrastive study. *Languages in Contrast*, 6(1), 109-149. doi: 10.1075/lic.6.1.05xia
- Xiao, R., & Yue, M. (2009). Using corpora in translation studies: The state of the art. In P. Baker (Ed.), *Contemporary Corpus Linguistics* (pp. 237-262). London: Continuum.
- Zanettin, F. (2013). Corpus methods for descriptive translation studies. *Procedia - Social and Behavioral Sciences*, 95, 20-32. doi: 10.1016/j.sbspro.2013.10.618

References in Chinese

- 曹先擢 (2005). 现代汉语词典(第五版). 北京: 商务印书馆.
- Cao, X. (2005). *Xiandai hanyu cidian [Modern Chinese Dictionary (5th ed.)]*. Beijing: The Commercial Press.
- 陈国鹏、周晓麟 (1989). 儿童对因果关系词的理解. 《上海教育科研》, 89 (4): 43-44.
- Chen, G., & Zhou, X. (1989). Ertong dui yinguo guanxici de lijie [Children's understanding of causal conjunctives]. *Shanghai Jiaoyu Keyan [Journal of Shanghai Educational Research]*, 89(4), 43-44.
- 董娟娟 (2010). 论汉语第三人称代词的欧化现象. 《云南师范大学学报》, 1: 68-73.
- Dong, J. (2010). Lun hanyu disan rencheng daici de ouhua xianxiang [The Europeanisation of Chinese third person reference]. *Yunnan shifan daxue xuebao [Journal of Yunnan Normal University]*, 8(1), 68-73.
- 高再兰 (2013). 前、后置“因为”的隐现及功能差异. 《汉语学报》, 2013 (2): 57-65.
- Gao, Z. (2013). Qian/hou zhi “yinwei” de yinxian ji gongneng chayi [On the omission of the preposed and postposed “yinwei (因为)” and the functional differences between them]. *Hanyu xuebao [Chinese Linguistics]*, 2, 56-65.
- 胡显耀 (2006). 《当代汉语翻译小说规范的语料库研究》. 华东师范大学博士论文.
- Hu, X. (2006). *Dangdai fanyi xiaoshuo de guifan yuliaoku yanjiu [A corpus-based study on the translational norms of contemporary Chinese translated fiction (Unpublished PhD dissertation)]*. East China Normal University, Shanghai.
- 胡显耀 (2007). 基于语料库的汉语翻译小说词语特征研究. 《外语教学与研究》, 2007 (3): 214-220.
- Hu, X. (2007). Jiyu yuliaoku de hanyu fanyi xiaoshuo ciyu tezheng yanjiu [A corpus-based study on the lexical features of Chinese translated fiction]. *Waiyu jiaoxue yu yanjiu [Foreign Language Teaching and Research]*, 39(3), 214-220.
- 胡显耀、曾佳 (2009). 对翻译小说语法标记外显化的语料库研究. 《外语研究》, 2009 (5): 72-79.
- Hu, X., & Zeng, J. (2009). Dui fanyi xiaoshuo yufa biaoji wai xian hua de yuliaoku yanjiu [A corpus-based study on the explicitation of grammatical markers in translated novels]. *Waiyu yanjiu [Foreign Languages Research]*, 117(5), 72-79.
- 黄伯荣、廖序东 (2002). 现代汉语 (第三版). 北京: 高等教育出版社.
- Huang, B., & Liao, X. (2002). *Xiandai hanyu [Modern Chinese] (3rd ed.)*. Beijing: Higher Education Press.
- 黄立波 (2007). 《基于汉英/英汉平行语料库的翻译共性》. 上海: 复旦大学出版社.

- Huang, L. (2007). *Jiyu hanying / yinghan pingxing yuliaoku de fanyi gongxing* [Translation universals research based on Chinese-English/English-Chinese parallel corpus]. Shanghai: Fudan University Press.
- 黄立波 (2008). 英汉翻译中人称代词主语的显化. 《外语教学与研究》, (6): 454-459.
- Huang, L. (2008). Yinghan fanyi zhong rencheng daici zhuyu de xian hua [Explication of personal pronoun subjects in English-Chinese translation: A corpus-based investigation]. *Waiyu jiaoxue yu yanjiu* [Foreign Language Teaching and Research], 40(6), 454-459.
- 蒋甜雨 (2016). 《基于语料库的 3-6 岁儿童文学英汉翻译常规化研究》. 华东理工大学硕士论文.
- Jiang, T. (2016). *Jiyu yuliaoku de 3-6 sui ertong wenxue yinghan fanyi changguihua yanjiu* [A corpus-based study on the normalisation of Chinese translated young children's literature for the 3-6-year-old (Unpublished master's thesis)]. East China University of Science and Technology, Shanghai.
- 柯飞 (2005). 翻译中的隐和显. 《外语教学与研究》, 2005 (4): 303-307.
- Ke, F. (2005). Fanyi zhong de yin he xian [The explication and implicitation of translation]. *Waiyu jiaoxue yu yanjiu* [Foreign Language Teaching and Research], 4, 303-307.
- 李宏顺 (2014). 国内外儿童文学翻译研究及展望. 《外国语》, 37 (5): 64-72.
- Li, H. (2014). Guoneiwai ertong wenxue fanyi yanjiu ji zhanwang [Research of children's literature translation and its future trends]. *Waiguoyu* [Journal of Foreign Languages], 37(5), 64-72.
- 李忠美、陈海庆 (2009). 指示语“我们”与“咱们”的异同及其语用含义探析. 《现代语文》, 5: 33-35.
- Li, Z., & Chen, H. (2009). Zhishiyu “women” yu “zanmen” de yitong jiqi yuyong hanyi tanxi [Investigating the differences and similarities between “women (我们)” and “zanmen (咱们)”]. *Xiandai yuwen* [Modern Chinese], 5, 33-35
- 吕叔湘 (1999). 现代汉语八百词. 北京: 商务印书馆.
- Lü, S. (1999). *Xindai hanyu babai ci* [Modern Chinese eight hundred words]. Beijing: The Commercial Press.
- 梅林 (2015). 《基于自建小型译文语料库的汉译少年文学语言特征研究》. 安徽大学硕士论文.
- Mei, L. (2015). *Jiyu zijian xiaoxing yiwen yuliaoku de hanyi shaonian wenxue yuyan tezheng yanjiu* [A corpus-based study on the linguistic features of the translated Chinese children's literature (for children aged between 12 and 18) (Unpublished master's thesis)]. Anhui University, Anhui.
- 潘文国 (2010). 汉英语言对比概论. 北京: 商务印书馆.

- Pan, W. (2010). *Hanying yuyan duibi gailun [The introduction to the contrastive studies of English and Chinese]*. Beijing: The Commercial Press.
- 皮亚杰 (1980). 儿童的语言与思维 (傅统先译). 北京: 文化教育出版社.
- Pi, Y. (1980). *Ertong de yuyan yu siwei [Children's language and thinking (translated by Fu Tongxian)]*. Beijing: People's Education Press.
- 孙丹萍 (2013). 从因果关系连词使用对比看英汉的形合与意合. 《赤峰学院学报 (汉文哲学社会科学版)》, 2013 (9): 204-207.
- Sun, D. (2013). Cong yinguo guanxi lianci shiyong duibi kan yinghan de xinghe yu yihe [The differences between English and Chinese in the use of causal conjunctions]. *Chifeng xueyuan xuebao (hanwen zhexue shehui kexue ban) [Journal of Chifeng University (Soc.Sci)]*, 34(9), 204-207.
- 王克非、胡显耀 (2010). 汉语文学翻译中人称代词的显化和变异. 《中国外语》, 2010 (4): 16-21.
- Wang, K., & Hu, X. (2010). Hanyu wenxue fanyi zhong rencheng daici de xianhua he bianyi [The explicitation and deviation of personal pronouns in Chinese literary translation]. *Zhongguo waiyu [Foreign Languages in China]*, 4, 16-21.
- 王丽丽 (2013). 《基于语料库的英译汉儿童文学显化研究》. 曲阜师范大学硕士论文.
- Wang, L. (2013). *Jiyu yuliaoku de yingyihan ertong wenxue xianhua yanjiu [A corpus-based study on explicitation in E-C translation of children's literature (Unpublished master's thesis)]*. Qufu Normal University, Qufu.
- 王良杰 (2006). 谈“吗”和“呢”的用法. 《唐山师范学院学报》, 2006 (3): 33-36.
- Wang, L. (2006). Tan ma he ne de yongfa [The usage of “吗” and “呢”]. *Tangshan shifan xueyuan xuebao [Journal of Tangshan Teachers College]*, 28(3), 33-36.
- 王力 (1984). 《中国语法理论》. 北京: 中华书局.
- Wang, L. (1984). *Zhongguo yafa lilun [Chinese grammar theory]*. Beijing: Zhonghua Book Company.
- 王伟 (2007). 英语中因果关系的表达. 《焦作师范高等专科学校学报》, 23 (3): 10-12.
- Wang, W. (2007). Yingyu zhong yinguo guanxi de biaoda [Causality in English]. *Jiaozuo shifan gaodeng zhuanke xuexiao xuebao [Journal of Jiaozuo Teachers College]*, 23(3), 10-12.
- 吴光华. 英汉大辞典 (第三版). 上海: 上海译文出版社.
- Wu, G. (2010). *Yinghan da cidian [The Chinese-English Dictionary (3rd ed.)]*. Shanghai: Shanghai Translation Publishing House.
- 许文胜、张柏然 (2006). 基于英汉名著语料库的因果关系连词对比研究. 《外语教学与研究 (外国语文双月刊)》, 2006 (4): 292-296.

- Xu, W. & Zhang, B. (2006). Jiyu yinghan mingzhu yuliaoku de yinguo guanxi lianxi duibi yanjiu [Corpus-based contrastive studies on the causal conjunctions in English-Chinese classics]. *Waiyu jiaoxue yu yanjiu (waiguo yuwen shuangyuekan) [Foreign Language Teaching and Research (bimonthly)]*, 38(4), 292-296.
- 于洋 (2014). 《基于语料库的 3-6 岁儿童绘本翻译语言研究》. 吉林大学硕士学位论文.
- Yu, Y. (2014). *Jiyu yuliaoku de 3-6 sui ertong huiben fanyi yuyan yanjiu [A corpus-based research of language use in 3-6 years old translated Chinese children's picture books (Unpublished master's thesis)]*. Jilin University, Jilin.
- 臧光亚 (2010). 《基于语料库的儿童文学翻译语言研究》. 曲阜师范大学硕士论文.
- Zang, G. (2010). *Jiyu yuliaoku de ertong wenxue fanyi yuyan yanjiu [A corpus-based study of language use in translated Chinese children's literature (Unpublished master's thesis)]*. Qufu Normal University, Qufu.
- 赵宏、邵志洪 (2002). 英汉第三人称代词语篇照应功能对比研究. 《外语教学与研究》, 3: 174-179.
- Zhao, H., & Shao, Z. (2002). Yinghan disan rencheng daici yupian zhaoying gongneng duibi yanjiu [A comparsion of third person pronouns in English and Chinese]. *Waiyu jiaoxue yu yanjiu [Foreign Language Teaching and Research]*, 34(3), 174-179.
- 赵应吉 (1995). 论“因为”一词的英译法. 《重庆科技学院学报》, 10 (4): 82-86.
- Zhao, Y. (1995). Lun “yinwei” yici de yingyifa [On the conversion of Chinese “因为” into English]. *Chongqing gangtie gaodeng zhuanke xuexiao xuebao [Journal of Chongqing University of Science and Technology]*, 10(4), 82-86.
- 朱曼殊、华红琴 (1992). 儿童对因果复句的理解. 《心理科学》, 1992 (8): 1-7.
- Zhu, M., & Hua, H. (1992). Ertong dui yinguo fuju de lijie [Children's understanding of the compound sentence of causality]. *Xinli kexue [Psychology Science]*, 8, 1-7.

Appendices

Appendix 1: Translated books for TCCLC

	Translated title	Title	Translator	Gender	Publish date	Publisher	Author	Gender	Originally published	Nationality
1	秘密花园	The Secret Garden	环宇宏基 Huanyuhongji	-	2011	延边人民出版社 Yanbian People Publishing House	Frances Hodgson Burnett	F	1911	American
2	蓝色的海豚岛	Island of the Blue Dolphins	傅定邦 Fu Dingbang	M	2014-10	新蕾出版社 Xinlei Publishing House	Scott O'Dell	M	1960	American
3	夏洛的网	Charlotte's Web	任溶溶 Ren Rongrong	M	2014-8	上海译文出版社 Shanghai Translation Publishing House	E. B. White	M	1952	American
4	精灵鼠小弟	Stuart Little	任溶溶 Ren Rongrong	M	2016-1	上海译文出版社 Shanghai Translation Publishing House	E. B. White	M	1945	American
5	绿野仙踪	The Wonderful Wizard of Oz	陈伯吹 Chen Bochui	M	2016-5	西安出版社 Xi'an Publishing House	L. Frank Baum	M	1900	American
6	长腿叔叔	Daddy-Long-Legs	董燕 Dong Yan	F	2015-3	浙江少年儿童出版社 Zhejiang Juvenile & Children's Publishing House	Jean Webster	F	1912	American
7	永远的狄家/不老泉	Tuck Everlasting	陈政一 Chen Zhengyi	N/A	N/A	N/A	Natalie Babbitt	F	1975	American
8	哈克贝利·芬历险记	Adventures of Huckleberry Finn	N/A	N/A	2004-12	西藏人民出版社 The Tibet people Publishing House	Mark Twain	M	1884	American
9	爱德华的奇妙之旅	The Miraculous Journey of Edward Tulane	王昕若 Wang Xinruo	F	2014-2	新蕾出版社 Xinlei Publishing House	Kate DiCamillo	F	2006	American
10	时间的皱纹	A Winkle in Time	廖丽 Liaoli	F	2007-6	吉林文史出版社 Jinlin literature & History Publishing House	Madeleine L'Engle	F	1963	American
11	小淘气	Rascal	吴淑玲 Wu Shuling	F	1998	中国少年儿童出版社 China Juvenile & Children's Books Publishing House	Sterling Noth	M	1963	American
12	汤姆的午夜花园/大座钟的秘密	Tom's Midnight Garden	马爱农 Ma Ainong	F	2006	人民文学出版社 People's Literature Publishing House	Philippa Pearce	F	1963	English
13	爱丽丝漫游奇境记	Alice's Adventures in Wonderland	管绍淳 Guan Shaochun	M	2014-1	华东师范大学出版社 East China Normal University Press	Lewis Carroll	M	1865	English
14	水孩子	The Water Babies	周煦良 Zhou Xuliang	M	2004-1	中国少年儿童出版社 China Juvenile & Children's Books Publishing House	Charles Kingsley	M	1863	English
15	小飞侠彼得·潘	Peter Pan	杨静远 Yang Jingyuan	F	2013-1	中国画报出版社 China Pictorial Publishing House	J. M. Barrie	M	1904	English
16	格列佛游记	Gulliver's Travels	蒋剑锋 Jiang Jianfeng	M	N/A	N/A	Jonathan Swift	M	1726	English
17	柳林风声	The Wind in the Willows	杨静远 Yang Jingyuan	F	2016-4	四川少年儿童出版社 Sichuan Juvenile & Children's Books Publishing House	Kenneth Grahame	M	1908	English
18	随风而来的玛丽·波平斯阿姨	Marry Poppins	任溶溶 Ren Rongrong	M	2012-5	明天出版社 Tomorrow Publishing House	P. L. Travers	F	1934	English
19	天蓝色的彼岸	The Great Blue Yonder	张雪松 Zhang Xuesong	M	2005-4	新世界出版社 New World Publishing House	Alex Shearer	M	2002	English
20	宝岛/金银岛	Treasure Island	路旦俊 Lu Danjun	M	2005-1	中国工人出版社 China Workers Publishing House	Robert Louis Stevenson	M	1883	English
21	大空人遇险记	Down to Earth	任溶溶 Ren Rongrong	M	1999-4	中国少年儿童出版社 China Juvenile & Children's Books Publishing House	Patricia Wrightson	F	1965	Australian
22	红头发安妮/绿山墙的安妮	Anne of Green Gables	吴方 Wu Fang	N/A	1999-2	南海出版公司 Hainan Publishing House	Lucy Maud Montgomery	F	1908	Canadian

M: male; F: female; N/A: not available; the English names of the Chinese publishers are either from the Internet or translated by the author of the thesis.

Appendix 2: Non-translated books for NCCLC

	Title	English translation	Author	Gender	Publish date	Publisher
1	狼王梦	The Dream of the King Wolf	沈石溪 Shen Shixi	M	2013-1	浙江少年儿童出版社 Zhejiang Juvenile & Children's Publishing House
2	男生贾里女生贾梅	Boy JiaLi Girl Jia Mei	秦文君 Qin Wenjun	F	2014-9	中国少年儿童出版社 China Juvenile & Children's Publishing House
3	青铜葵花	Qingtong Kuihua (two names)	曹文轩 Cao Wenxuan	M	2008-5	江苏人民出版社 Jiangsu People Publishing House
4	我的妈妈是精灵	My Mum is An Fairy	陈丹燕 Chen Danyan	F	2014-10	福建少年儿童出版社 Fujian Juvenile & Children's Publishing House
5	笑猫日记-寻找黑骑士	Smiling Cat's Diary-Looking For Black Knight	杨红樱 Yang Hongying	F	2008-1	明天出版社 Tomorrow Publishing House
6	笑猫日记-绿狗山庄	Smiling Cat's Diary-Green Dog Mountain Villa	杨红樱 Yang Hongying	F	2013-4	明天出版社 Tomorrow Publishing House
7	西游记-童话大王讲经典	Journey to the West	吴承恩 (著) / 郑渊洁 (改编) Zheng Yuanjie	M	2011-8	二十一世纪出版社 21st Century Publishing House
8	大林和小林	Big Lin and Little Lin	张天翼 Zhang Tianyi	M	2012-8	北方妇女儿童出版社 Northern China Women & Children Publishing House
9	宝葫芦的秘密	The Secret of the Calabash	张天翼 Zhang Tianyi	M	2015-5	安徽教育出版社 Anhui Education Publishing House
10	小灵通漫游未来	Xiao Lingtong Wonder in the future	叶永烈 Ye Yonglie	M	2016-4	长江少年儿童出版社 Changjiang Juvenile & Children's Publishing House
11	“下次开船”港	"Next Time Depart" Bay	严文井 Yan Wenjing	M	2013-8	陕西人民教育出版社
12	魔法听诊器	The Magic Stethoscopy	商晓娜 Shang Xiaona	F	2010-2	福建少年儿童出版社 Fujian Juvenile & Children's Publishing House
13	魔法学校·小女巫	Magic School-A Little Witch	葛竞 Ge Jing	F	2015-10	春风文艺出版社 Chunfeng Literature & Art Publishing House
14	白狐迪拉与月亮石	White Fox Dila and Moon Stone	陈佳同 Chen Jiatong	M	2014-9	人民文学出版社 People's Literature Publishing House
15	小老虎历险记	The Adventure of A Little Tiger	汤素兰 Tang sulan	F	2013-4	天天出版社 Daylight Publishing House
16	飞向人马座	Flying to the Sagittarius	郑文光 Zheng Wenguang	M	2006-9	湖北少年儿童出版社 Hubei Juvenile & Children's Publishing House
17	装在口袋里的爸爸·后悔药	Dad in the Pocket-Regreting Medicine	杨鹏 Yangpeng	M	2013-6	二十一世纪出版社 21st Century Publishing House
18	没有风的扇子	A Fan without Wind	孙幼军 Sun Youjun	M	2015-6	江苏凤凰少年儿童出版社 Jiangsu Phoenix Juvenile & Children's Publishing House
19	金猫历险记	The Adventure of A Golden Cat	孙大文 Sun Dawen	M	N/A	N/A
20	中国五十年儿童文学名家作品选	The collection of children's literature written by great writers in the past 50 years	N/A	N/A	N/A	N/A

M: male; F: female; N/A: not available; the English book titles are translated by the author of the thesis; the English names of the Chinese publishers are either from the Internet or translated by the author of the thesis.

Appendix 3: Quantile-quantile plots to assess normality

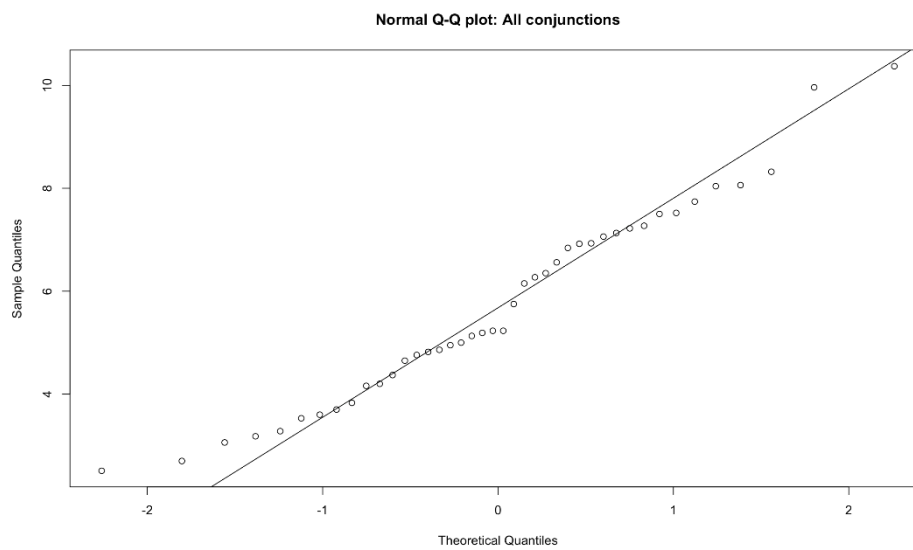


Figure 1: Quantile-quantile plot to assess normality: all conjunctions

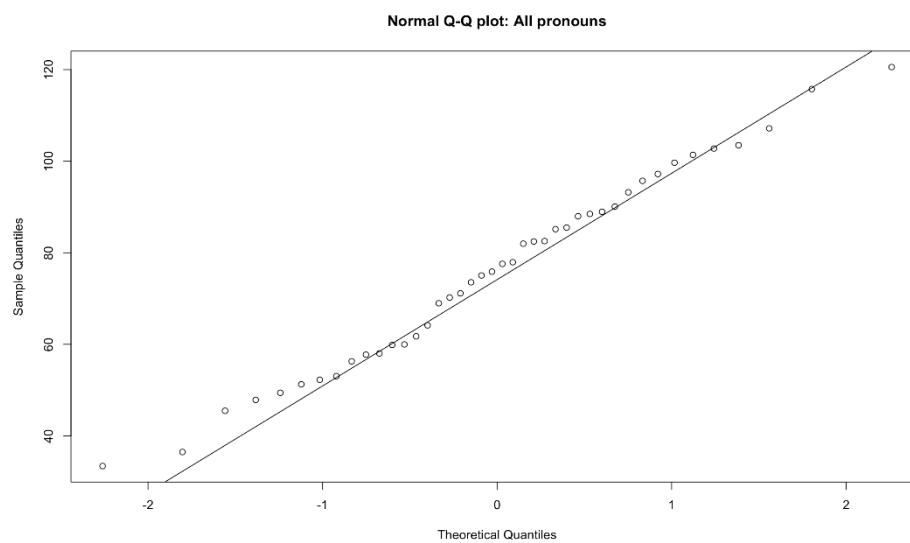


Figure 2: Quantile-quantile plot to assess normality: all pronouns

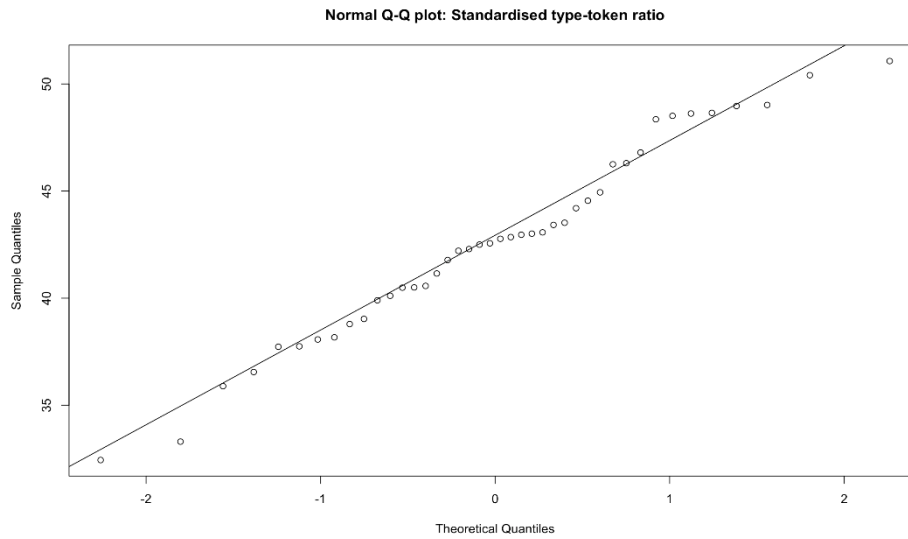


Figure 3: Quantile-quantile plot to assess normality: STTR

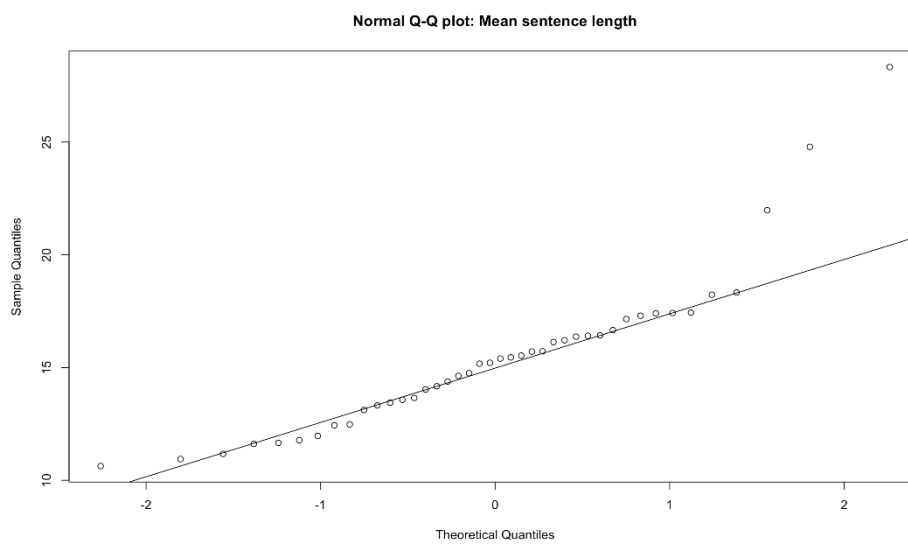


Figure 4: Quantile-quantile plot to assess normality: mean sentence length

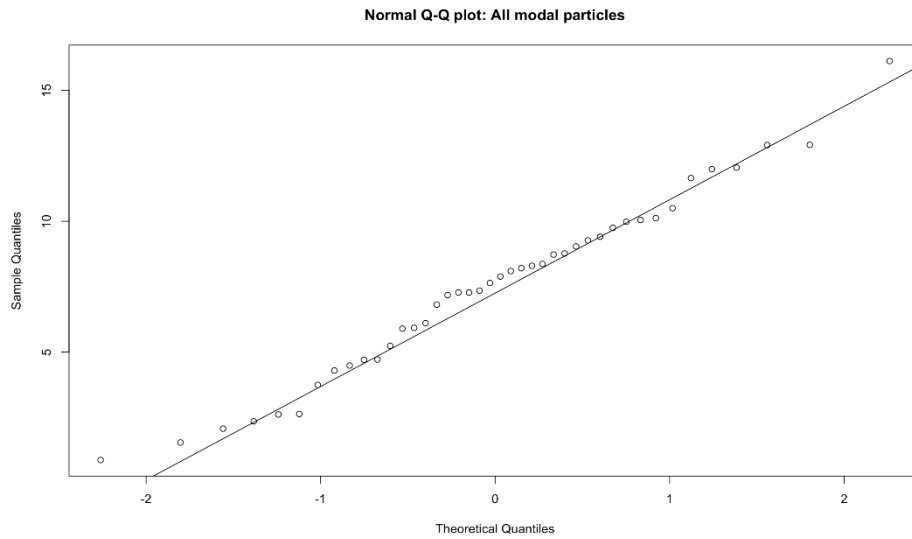


Figure 5: Quantile-quantile plot to assess normality: all modal particles

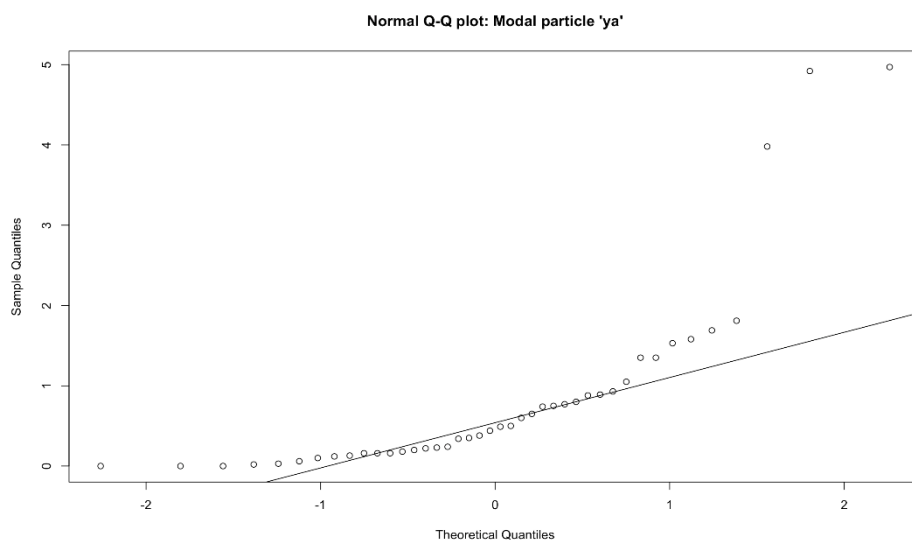


Figure 6: Quantile-quantile plot to assess normality: modal particle 呀 *ya*