

Passive, Deontic Modality and Cohesive  
Conjunction in English-to-Thai Legislative  
Translation: A Corpus-Based Study

Mali Satthachai

PhD

2019



# Passive, Deontic Modality and Cohesive Conjunction in English-to-Thai Legislative Translation: A Corpus-Based Study

**Mali Sattachai**

MA, BA

Thesis submitted in fulfilment of the requirements for the degree  
of

Doctor of Philosophy (PhD)

Dublin City University

School of Applied Language and Intercultural Studies

Supervisor: Professor Dorothy Kenny

March 2019

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy (PhD) is entirely my own work, and that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: \_\_\_\_\_

(Mali Satthachai)

ID No.: 14211473

Date: \_\_\_\_\_

## **Acknowledgements**

First and foremost, I would like to express my deepest gratitude to Professor Dorothy Kenny, my thesis supervisor for everything she has done for me. Her valuable support and assistance has exceeded the responsibility of a thesis supervisor. Her professional guidance has been a source of enlightenment and inspiration. Without her contribution, this thesis would not have been possible.

I appreciate the support and facilities provided by staff at School of Applied Language and Intercultural Studies, Faculty of Humanities, Dublin City University and particularly appreciate the help given by the former head of school, Professor Dorothy Kenny, and the present head of school, Professor Françoise Blin, and also the school secretaries, Caroline Whitston and Paula Smith.

I wish to thank my family for their financial and moral support. I also would like to offer my special thanks to my friends George and Raweewan Smullen and John and Wilaiporn Perkins for their encouragement and kind support throughout my study in Dublin.

Finally, I wish to thank Chiang Mai University for their financial support from September 2014 to September 2018.

## **Publications and Conference Presentations Based on this Research**

### **Publications**

Kenny, Dorothy and Mali Sathachai (2018, in press) ‘Explicitation, Unique Items and the Translation of English Passives in Thai Legal Texts’, *Meta* 63(3).

Sathachai, Mali and Dorothy Kenny (2019, in press) ‘Deontic Modality in English-Thai Legislative Translation’, *Translation Spaces* 8(1).

### **Conference Papers**

Kenny, Dorothy and Mali Sathachai (2017) ‘Translation Features in an English-Thai Legal Translation Corpus’, Paper presented at the *1<sup>st</sup> World Congress on Translation Studies*, 10-14 April 2017, Université Paris-Nanterre, France.

Sathachai, Mali and Dorothy Kenny (2018) ‘Deontic Modality in English-Thai Legal Translation Corpora’, Paper presented at the *Transius Conference*, 18-20 June 2018, University of Geneva, Switzerland.

## Table of Contents

<i>Acknowledgements</i> .....	ii
<i>Publications and Conference Presentations Based on this Research</i> .....	iii
<i>List of Tables and Figures</i> .....	ix
<i>Abstract</i> .....	xi
<b>Introduction</b> .....	<b>1</b>
Thesis Aims .....	2
Data Sources and Basic Methodology .....	3
Thesis Structure .....	3
The Thai Language .....	4
<i>General Observations</i> .....	5
<i>Thai Lexis</i> .....	6
<i>Thai Syntax</i> .....	8
Transcription and Data Presentation .....	10
<b>Chapter 1: Legal Language and Legal Translation</b> .....	<b>13</b>
Introduction .....	13
Legal Language .....	13
Typologies of Legal Language .....	14
Treaties .....	16
Legal English .....	18
<i>Main Features of Legal English</i> .....	18
<i>Use of Legal English</i> .....	19
Legal Thai .....	19
<i>Thai Legal Periods</i> .....	20
<i>Foundation of Modern Thai Law</i> .....	21
<i>Features of Modern Legal Thai</i> .....	22
Legal Translation .....	23
<i>Typologies of Legal Translation</i> .....	23
<i>Characteristics of Legal Translation as a Source of Difficulty</i> .....	24
System-Bound Nature .....	25
Linguistic Peculiarities .....	26
Translating Law as a Problem-Solving Process .....	27
<i>Equivalence</i> .....	28
Legal Equivalence .....	29
Equal Intent .....	30

Fidelity.....	31
Creativity .....	32
Interpretation.....	33
Conclusions .....	33
<b>Chapter 2: Corpus-Based Descriptive Translation Studies and Translation Features .....</b>	<b>35</b>
Introduction .....	35
Explicitation .....	37
<i>Definition and Criticism</i> .....	37
<i>Typologies of Explicitation</i> .....	39
<i>Major Studies and Criticism</i> .....	40
<i>Studies of Explicitation in Asian Contexts</i> .....	44
<i>Different Approaches to Explicitation</i> .....	46
Explicitation as a Feature of Language Mediation.....	46
Explicitation in Relation to Readers .....	47
Explicitation and Risk.....	48
Unique Items Hypothesis .....	49
<i>Definition and Criticism</i> .....	49
<i>Major Studies</i> .....	51
Source-Language Shining Through .....	53
<i>Definition</i> .....	53
<i>Major Studies</i> .....	53
Studies of Translation Features in Legal Texts .....	55
Why Study Translation Features? .....	58
Conclusions .....	60
<b>Chapter 3: Methodology .....</b>	<b>61</b>
Introduction .....	61
Corpus Resources in Thai .....	61
Corpora in Use .....	63
<i>Corpus Design for Special Purpose Text</i> .....	63
<i>Parallel Corpus Design</i> .....	64
Language Direction.....	65
Size .....	65
Genre, Authorship and Copyright .....	66
Text Extent.....	66
Authenticity .....	66



Parallel Corpus Text Preparation .....	68
<i>Monolingual Legal Thai Comparable Corpus</i> .....	69
Size .....	69
Genre, Authorship, Copyright and Text Status .....	69
Comparability .....	71
Text Preparation.....	72
<i>Software Used: Sketch Engine vs ParaConc</i> .....	72
<i>Research Questions</i> .....	74
<i>Operationalization of Key Concepts</i> .....	75
Explicitation.....	75
Unique Items.....	76
Source language shining through .....	77
<i>Selection of Features and Methods for Analysis</i> .....	77
Teich's (2003) Methodology .....	77
Becher's (2011) methodology.....	78
Biel's (2014) Methodology .....	79
Methods Used in this Study .....	80
Conclusions .....	83
<b>Chapter 4: Passive Voice.....</b>	<b>84</b>
Introduction .....	84
The Passive Voice in Thai.....	84
<i>Passive with Passive Markers / 'Overt' Passive</i> .....	85
ถูก /thuk/ and โดน /don/ passives .....	85
ได้รับ /dai/+/rup/ passive.....	87
<i>Passive without Passive Markers / 'Covert' Passive</i> .....	88
Research Questions .....	90
Research Procedure .....	90
<i>English-Thai Passives in Bilingual Parallel Corpus</i> .....	92
<i>Translated vs Non-Translated Thai</i> .....	100
Discussion .....	101
<i>Relation between Source and Target Texts</i> .....	101
<i>Textual Fit</i> .....	103
Conclusions .....	103
<b>Chapter 5: Deontic Modality .....</b>	<b>105</b>
Introduction .....	105
Deontic Modality in English and Thai .....	106

Modal Strength.....	107
Modality and Modal Strength in Translation .....	108
Research Questions .....	110
Research Procedure .....	110
Results and Analysis .....	111
<i>Deontic modal verbs in the English source texts</i> .....	111
<i>Thai translations of English deontic modal verbs</i> .....	113
<i>Deontic Modals in Monolingual Thai and Thai Translation</i> .....	119
Discussion .....	121
<i>Source-Language Modals</i> .....	121
<i>Relations between Source and Target Texts</i> .....	122
<i>Textual Fit</i> .....	125
Conclusions .....	127
<b>Chapter 6: Cohesive conjunction .....</b>	<b>128</b>
Introduction .....	128
Cohesion.....	128
Conjunction.....	128
Thai Conjunction.....	130
Cohesive Conjunction Investigated in this Research .....	131
Research Questions .....	132
Research Procedure .....	132
Results and Analysis .....	135
<i>Cohesive Conjunction in the English-Thai Parallel Corpus</i> .....	135
Additive Cohesive Conjunction .....	135
Adversative Cohesive Conjunction .....	139
Temporal Cohesive Conjunction .....	143
Causal Cohesive Conjunction .....	149
Conditional Cohesive Conjunction .....	151
Summary of Findings on Cohesive Conjunction in Source and Target Texts .....	158
<i>Cohesive Conjunction in Legal Thai Monolingual and Thai Translated Texts</i> .....	159
Additive Cohesive Conjunction .....	159
Adversative Cohesive Conjunction .....	161
Temporal Cohesive Conjunction .....	162
Causal Cohesive Conjunction .....	163
Conditional Cohesive Conjunction .....	164

<i>Cohesive Conjunctions in English vs Thai Translated Texts vs Thai Non-Translated Texts</i>	166
Discussion .....	167
Conclusions .....	168
<b>Chapter 7: Conclusions .....</b>	<b>170</b>
Theoretical Framework and Methodology .....	170
Recapitulation of Research.....	171
Contribution of the Research.....	173
Limitations of the Research and Avenues for Further Study .....	174
<b>References.....</b>	<b>176</b>
<b>Appendices .....</b>	<b>193</b>

## List of Tables and Figures

### Tables

3.1	English-Thai parallel corpus design	68
3.2	Thai monolingual corpus design	70
4.1	Three main techniques employed by Thai translators when translating English passives	93
4.2	Occurrences of passive markers /thuk/ and /dai/+/rup/ in Thai translation and comparable non-translated Thai	100
5.1	Frequency of modal verbs in the English source texts	112
5.2	Deontic obligation and permission modal verbs	113
5.3	Thai translations of deontic modal ‘shall’	113
5.4	Thai translations of deontic modal ‘should’	115
5.5	Thai translations of deontic modal ‘will’	116
5.6	Thai translations of deontic modal ‘may’	117
5.7	Thai translations of deontic modal ‘can’	119
5.8	Thai deontic modals in translation and monolingual sub-corpora	120
6.1	Candidate instances of cohesive conjunction	135
6.2	Additive cohesive conjunction	139
6.3	Adversative cohesive conjunction	143
6.4	Temporal cohesive conjunction	148
6.5	Causal cohesive conjunction	151
6.6	Conditional cohesive conjunction	157
6.7	Additive conjunction in the Thai monolingual and translation corpora	161
6.8	Adversative conjunction in the Thai monolingual and translation corpora	162
6.9	Temporal conjunction in the Thai monolingual and translation corpora	163
6.10	Causal conjunction in the Thai monolingual	

	and translation corpora	164
6.11	Conditional conjunction in the Thai monolingual and translation corpora	165
6.12	Comparison of frequency of cohesive conjunctions in English and Thai monolingual corpora	166

## Figure

4.1	Frequency per million words of passive markers /thuk/ and /dai/+/rup/ in Thai translated and Thai non-translated legal corpora	100
-----	-----------------------------------------------------------------------------------------------------------------------------------	-----

## **Abstract**

### **Passive, Deontic Modality and Cohesive Conjunction in English-to-Thai Legislative Translation: A Corpus-Based Study**

Mali Satthachai

Scholarly interest in legislative translation has grown substantially over recent decades, with corpus-based approaches contributing much to our understanding of the relationship between translated legislation and source texts, on the one hand, and translated and non-translated legislative texts in the target language, on the other. To date, however, most studies have been conducted on European languages. This study represents a first attempt to use corpus techniques to explore legislative translation from English into Thai. Drawing on a purpose-built, 400,000-word, parallel corpus of international treaties translated from English into Thai, and a one million-word monolingual corpus of legislative texts originally written in Thai, both of which are managed using the Sketch Engine platform, we investigate how selected linguistic features—including instances of passive voice, deontic modality and cohesive conjunction—are translated into Thai. We analyze the inter-linguistic and intra-linguistic differences we find in the light of the possible adoption of ‘plain writing’ principles (Williams 2009) in Thai legislation, the legal authenticity of the Thai non-translated texts, previously posited ‘features of translation’ including Teich’s (2003) notion of source texts ‘shining through’, Becher’s (2010) approach to explicitation and the presence of ‘unique items’ (Tirkkonen-Condit 2002), and drawing on Biel’s (2014) concept of ‘textual fit’.

## Introduction

Despite growing interest in translation studies in Thailand, there is still relatively little published on contemporary Thai translation practices in general (Wakabayashi 2014) and contemporary Thai legal translation in particular. There is good reason to believe, however, that Thai legal translation might be a ‘rich field’ (ibid.) for translation studies scholars, given its atypical development: Siam, as contemporary Thailand was previously called, is often seen as exceptional in that although it was a “victim of European imperial aggression” (Loos 1999: 2), it was not actually colonized. At the same time, it harboured imperial ambitions of its own, and established a legal regime that mirrored aspects of the legal systems of European colonizing powers (ibid.). Of particular interest here is that fact that when legal reform was undertaken between the 1890s and the 1930s, partly in a bid to keep European colonial powers at bay, it was carried out by a cosmopolitan group of Siamese and foreign experts, who used continental (code) law as a prototype, but who drafted new laws in ‘functional’ English (Kraivixien 2012), and then translated them into Thai.

Even though modern Thai legislation is drafted directly in Thai, Kraivixien (2012: 43-50) argues that the functional English used during the earlier reform period continues to exert a positive influence over legal drafting in Thai, thanks to its valorization of concision and clarity, qualities that were subsequently brought over into legal Thai. The reform period could continue to exert influence over both Thai legal drafting and legal translation in other ways however: in translating from English, Thai legal reformers were expected to minimize the ‘inappropriate’ influence of the English language on Thai law. Passive voice, for example, which was used routinely in English legal texts, was frowned upon in Thai translation (ibid.: 171). Thai legislative translation thus opens up interesting vistas for translation scholars as it provides an intriguing context in which the source language, in our case English, not only exerts influence through the source texts themselves, but also through a specific tradition of legal drafting that, in turn, provided the model for legal drafting in Thai. At the same time, legal translators are warned against overusing structures that are typical of English. To date, to our knowledge, no research has looked at how this ambivalence towards English plays out in the linguistic make-up of legislative translations in Thai. The research presented in this thesis thus represents a first attempt to tackle this broad question. We are interested

in what typically happens in legislative translation from English into Thai, and whether the ensuing Thai translations are typical of legislative texts that originate in Thai. To use the dichotomy presented in Biel (2014), we are interested in both ‘equivalence’ relations between English legislative texts and their Thai translations, and the ‘textual fit’ of such Thai translations with other legislative texts in Thai.

Likewise, there is to date, again to our knowledge, almost no translation studies scholarship involving Thai that uses corpus linguistic methods. This might be understandable in light of the difficulties that the computational processing of written Thai can present (see below), but we would contend that there are some questions in translation studies that are best broached using corpus techniques, despite the considerable hurdles that may first have to be overcome. Such questions often involve linguistic features for which there may be hundreds, if not thousands, of instantiations in whatever body of text is under consideration, and for which computational assistance thus becomes important at least in quantitative research. Moreover, any linguistic research that is interested in what typically happens in translation can benefit from a corpus-based approach. Typicality is, after all, related to the frequency of occurrence of linguistic features, and this is best observed in corpora.

## **Thesis Aims**

Corpus-based translation studies have evolved since Baker’s (1993, 1995, 1996) seminal publications of the mid-1990s. More recent approaches stress the need for methodological rigour and statistical sophistication, and the benefits of integrating source-language influence, language-typological differences and genre effects into corpus-based studies of translation (see, for example, Becher 2011; De Sutter et al. 2012; Teich 2013; Biel 2014). In this thesis we respond to these broad trends by investigating how instances of specially selected features of legislative texts in English are typically translated into Thai, and how such translation fits, or not, with established patterns in the same genre in Thai. The features we focus on are passive voice, deontic modality and cohesive conjunction. Passive voice is selected because its use in English and Thai is expected to differ based on typological differences between the two languages. Such differences lie behind the existence in Thai of a category of passive marker that does not exist in English, and which thus can be seen as a ‘unique item’ (Tirkkonen-Condit 2002) in Thai (from the point of view of English at least). For reasons elaborated upon later in this thesis, instances of passive voice might also be



susceptible to what has become known as ‘explicitation’ (Becher 2011). Studies of passive voice thus allow us to investigate hypotheses from the translation studies literature that are related to unique items and explicitation. Deontic modality and cohesive conjunction—and specifically the subtype of cohesive conjunction known as ‘conditional conjunction’—on the other hand, are features that are closely associated with legislation (Biel 2014). Their importance to this legal genre is thus the prime motivation for their inclusion in our study, but they too offer opportunities to investigate the behaviour of unique items and the phenomenon of explicitation. They also allow us to track the influence of both the source language, which may or may not ‘shine through’ in translation (Teich 2003), and the target language genre, with which the translations’ textual fit can be gauged.

### **Data Sources and Basic Methodology**

The data investigated in this study come from a purpose-built 400,000-word parallel corpus of international treaties translated from English into Thai, and a one-million word monolingual corpus of legislation originally written in Thai, and which we refer to variously as our ‘monolingual Thai corpus’ or our ‘non-translated Thai corpus’. To our knowledge, our parallel corpus is the first such corpus to include Thai legislative texts. Our analysis is both quantitative and qualitative: we describe the distribution of selected features across our two corpora in terms of their frequency of occurrence, and we attempt to explain the inter-linguistic and intra-linguistic differences we find in the light of the possible adoption of ‘plain writing’ principles (Williams 2009) in Thai legislation, the legal authenticity of the Thai non-translated texts, previously posited ‘features of translation’ including Teich’s (2003) notion of source texts ‘shining through’, Becher’s (2010) approach to explicitation and the presence of ‘unique items’ (Tirkkonen-Condit 2002), and drawing on Biel’s (2014) concept of ‘textual fit’.

### **Thesis Structure**

The rest of this thesis is structured as follows. In the remaining part of this Introduction, we provide background information on the Thai language, and on how Thai data are presented in the thesis.

Chapter 1 then gives an overview of scholarship to date on English legal language, Thai legal language and legal translation in general.

Chapter 2 presents other elements of the theoretical framework used in the thesis, starting with the descriptive translation studies concept of ‘features of translation’ (formerly known as ‘translation universals’) and the major corpus-based studies in this area. Studies of translation features within the specific framework of legal translation are also addressed.

Chapter 3 presents our methodology, describing the corpora used in this research, and the procedure followed in extracting data from them. The chapter opens with some preliminary observations on the computational processing of Thai and the availability of corpora in the language pair English-Thai, factors that influence the methodology ultimately adopted in this research.

Chapters 4-6 present data from our corpora and our analysis of these data. In these chapters we use, in particular, the concepts of explicitation, the unique items hypothesis, and the notion of source-language shining through, as interpretive lenses through which we can examine the translation of passive voice (Chapter 4), deontic modality (Chapter 5) and cohesive conjunction (Chapter 6).

Chapter 7 presents the conclusions of the thesis as well as observations on the limitations of the study and plans for future research.

## **The Thai Language**

This section provides a brief account of basic grammatical phenomena in the Thai language, focusing on how Thai compares and contrasts with English. A description of general Thai is helpful at this point for a number of reasons. First, despite possessing some unique characteristics, legal Thai, which constitutes the main focus of this thesis, is still a subset of the Thai language in terms of its basic syntactic and morphological structure. Secondly, the literature on legal Thai is very limited and an understanding of how the Thai language operates in general might help give the reader a picture of how legal Thai operates in particular. Thirdly, Thai is considered an under-researched language within translation studies, and so it cannot be assumed that the reader has even a basic understanding of the structure of Thai, a situation that is unlikely to pertain in languages like English, for example. Fourthly, a basic understanding of how Thai is written will later help explain the particular challenges faced by the researcher when it comes to the automatic processing of Thai using corpus tools.

The section is organized as follows: the first part provides a general introduction to the Thai language and the second part reviews the fundamentals of Thai lexis and grammar with an emphasis on word classes and syntax. We stress here that the aim of this section is to provide a brief, general grammatical overview of Standard Thai. At other points in this thesis, and especially in Chapters 4, 5 and 6, we will require a somewhat deeper understanding of certain phenomena in Thai, and a more detailed account of those phenomena will be given then.

### ***General Observations***

Thailand is one of the major countries in Southeast Asia, with a population of around 65 million people as of 31 December 2014 (Department of Provincial Administration of Thailand, 2014). Thailand is a country of multilingualism with four major dialects and eighty distinct languages (Smalley 1994:1). Many Thais speak and understand more than one dialect or language (Iwasaki and Ingkaphirom 2005: 20). Multilingualism is not a cause of political instability however, since Standard Thai, the language under review in this thesis, is widely understood by speakers of other dialects and languages (ibid.). Standard Thai is, in fact, the national and official language of Thailand. Although it is the native language of just 19.5% of the population, it is used and understood widely throughout the country due to its use at school and in the mass media (Iwasaki and Ingkaphirom 2005:1).

Spoken Thai is a tonal language. There are five phonemic tones described as mid, low, falling, high and rising (Diller 2008: 33). The Thai writing system, according to popular belief, was devised by King Ramkhamheang of Sukhothai in 1283 (Brown 1985:5), and was modelled after the Khmer writing system, based on the system that originated in India (Iwasaki and Ingkaphirom 2005: 3). The modern Thai writing system uses 42 consonant letters and 32 vowel letters (Iwasaki and Ingkaphirom 2005:7). Other notable features of written Thai are that words are not delimited by spaces, that is, they are not ‘orthographic words’, and Thai sentences are, likewise, not delimited by either punctuation or spaces. This makes the computational processing of written Thai more challenging than it is for ‘white space’ languages.

The first Thai grammar book, *A Grammar of the Thai or Siamese Language*, was published in 1828 by James Low, an English diplomat-trader working in Thailand at that time. Prior to 1962, there were only a few books available in Thailand explaining

the Thai language in a systematic way, among which is Anuman Rajadhon's *The Thai Language* (1961). Knowledge of linguistics entered Thailand and thrived after 1962 through Thai graduates who had completed their studies in Western countries. Since then, the Thai language has been described within the framework of Western linguistics, and English-language linguistics in particular. On the one hand, this has been beneficial in that Thai has been explained in a more structural and systematic way. On the other hand, applying linguistic concepts that were first established in the West has created an unclear consensus on the Thai language, as will be further illustrated below.

### ***Thai Lexis***

There are five strata of vocabulary in Thai according to Iwasaki and Ingkaphirom (2005: 11): (i) indigenous vocabulary, (ii) Chinese loans, (iii) Khmer loans, (iv) Sanskrit and Pali loans, and (v) more recent loans from English and other languages. Thai mainly consists of monomorphemic words and multimorphemic loan words from Pali and Sanskrit however (Singhapreecha 2007:1497). Word formation involves compounding, reduplicating, and the affixation of loan words (Iwasaki and Ingkaphirom 2005: 11). Thai is an isolating language; unlike English, Thai lacks inflectional morphology to code grammatical information such as gender and number for nouns and verbs (Iwasaki and Ingkaphirom 2005:3).

There are no fixed rules on categorizing Thai words into classes; individual scholars sometimes have different views, and Thai words have been classified differently even by government organizations. For instance, the Thai Royal Institute divides Thai words into 8 classes (Rachabanditsathan 1999), while the Ministry of Education lists 12 word classes (Sriyapai 2013: 137). This could be the result of the previously mentioned application to Thai of linguistic categories that originate in the West. Moreover, distinguishing words on morphological grounds appears to be an impossible task in an isolating language like Thai (Iwasaki and Ingkaphirom 2005: 8). Thai words are also very fluid as they are still undergoing grammaticalization, the process by which content words such as noun and verbs become function words like prepositions and auxiliary verbs (ibid.).

As there is no absolute agreement on Thai word classes, this section classifies Thai words based on the categorisation provided by Iwasaki and Ingkaphirom (2005), who set up Thai word classes according to structural, semantic and functional criteria which

are comprehensible for general audiences and are consistent with our later explanations of Thai syntax. Iwasaki and Ingkaphirom (2005:8-11) divide Thai words into 14 classes as follows:

(1) Nouns: Nouns can function as a subject, the object of a verb or preposition, or a predicate after the copulas เป็น /pen/ and คือ /khue/. Thai nouns can be followed by various modifiers, such as adjectives, demonstratives, numerals, possessive phrases and relative clauses.

(2) Pronouns: Pronouns can be a subject, or the object of a verb or preposition. They can be followed by demonstratives but not adjectives, numerals, possessive phrases or relative clauses. A unique characteristic of Thai pronouns is their extreme inventory system of ranking. There are highly specialized terms such as อาตมา /aat-ta-ma/, which is used as a first person pronoun by a Buddhist monk when speaking with a lay person or low-ranking monk. There are also pronouns borrowed from other languages, such as ชู /yu/, a second-person pronoun borrowed from the English word ‘you’. In addition, the level of formality and referent’s sex can also be marked by the pronouns used.

(3) Demonstrative: Three demonstratives, นี่ /ni/, นั่น /nan/ and นั้น /nun/, are always used to provide details about proximity. The word นี่ /ni/ indicates close proximity to a speaker (equivalent to ‘this’ in English) while the words นั่น /nan/ and นั้น /nun/ indicate medial and distal proximity.

(4) Prepositions: Prepositions, as their name suggests, appear before nouns. Some prepositions are the grammaticalized form of verbs.

(5) Classifiers: Classifiers are a special category used when nouns are counted and modified. Classifiers are required for many common nouns in counting expressions and classifiers are also used to suggest definitiveness especially with deictics.

(6) Numerals: Numbers come after the nouns they modify.

(7) Verbs: Thai has intransitive, transitive and ditransitive verbs.

(8) Auxiliary verbs: These provide additional information for the main verbs such as aspect, modality, direction, and potentiality.

- (9) Negators: There are many negators in the Thai language but the common ones are *ไม่* /mai/ and *ไม่ใช่ว่า* /mai/+chai/.
- (10) Adjectives: Adjectives can be used either as modifiers or predicates. Predicate adjectives are always treated as stative verbs.
- (11) Adverbs: There are manner adverbs derived from adjectives and there are also genuine adverbs, which modify the whole sentence rather than a verb alone.
- (12) Linkers: There are several types of linkers. There are lexical linkers connecting words, clause linkers combining clauses and sentences and discourse linkers (markers) combining sets of propositions in discourse.
- (13) Particles: Five sub-types of particles are typically used, namely (i) speech-level marking particles, (ii) question particles, (iii) pragmatic particles, (iv) one linking particle and (v) one modal particle.
- (14) Exclamatives: Exclamatives are independent words that express a speaker's assessment of a situation or an emotion that s/he is experiencing.

### ***Thai Syntax***

As already indicated, approaches to Thai syntax have drawn on familiar Western classical grammatical categories, but they can also make use of some of Indic notions (Diller 2008: 37). In the following sub-sections, a brief description of nominal constructions, verbal constructions, clausal constructions and sentence types in Thai syntax will be given.

Nominal modification order in Thai strongly accords with the inherited Tai pattern which is [nominal head] + [modifier] (Diller 2008: 36). Noun phrases are head initial with the nominal head preceding modifiers such as adjectives, numerals and classifiers (Singhapreecha 2007: 1497). A nominal structure can range from simple to extended noun phrases, for example:

- (i) [noun+ adjective]: *หมาดำ* /ma/+dam/

: *หมา* / ma / (noun) = dog + *ดำ* /dam/ (adjective) = black

: English translation= black dog

(ii) [noun+adjective+numeral+classifier+demonstrative]

:หมาดำสามตัวนี้ /ma/+/dam/+/sam/+/tua/+/ni/

:หมา /ma/ (noun) = dog +ดำ /dam/ (adjective) = black +

สาม /sam/ (numeral) =three + ตัว /tua/ (classifier)+ นี้ /ni/ (demonstrative)= these

: English translation= these three black dogs

Nevertheless, there are a small number of cases where Indic order [modifier+noun] has an influence on Thai noun phrases. The pattern in which modifiers precede the nominal head can be seen in some commercial and institutional noun phrase names (Diller 2008: 36).

Thai verb phrases consist of a main verb alone or a main verb and one or more auxiliaries, and auxiliary verbs can be either pre-verbal or post-verbal (Iwasaki and Ingkaphirom 2005:14). Thai has no morphological markings for tense or agreement; however, Thai has a rich system of time adverbs and modal auxiliaries that can express tense, aspect and mood (Singhapreecha 2007: 1501). In addition, verbs can be simple or serial, where serial verb constructions contain a series of verbs arranged successively with the same subject (Singhapreecha 2007: 1501).

The negator 'ไม' /mai/ is commonly used with verbs. It can be placed before main verbs or certain auxiliaries, for example, 'ไมไป' /mai/+/pai/= 'ไม' /mai/ (not- negator) + ไป /pai/ (go- main verb) = not go.

Regarding the passive voice, the two pre-verbal auxiliaries ถูก /thuk/ and โดน/don/ are commonly used, for example, ถูกเขียน /thuk/+/khian/= ถูก /thuk/ (passive auxiliary) + เขียน /khian/ (write) = be written. The auxiliaries ถูก /thuk/ and โดน /don/ can be used in both adversative and neutral contexts (Singhapreecha 2007: 1509). According to Iwasaki and Ingkaphirom (2005: 313), adversity encoding is a major motivation for the use of passive in Thai.

As to commonly-cited typological parameters, Thai clauses most frequently display intransitive order [S+V-intrans] and basic transitive syntactic order [S+V-trans+O], while [V-intrans+S] also occurs in presentational or existential contexts (Diller 2008:

35).<sup>1</sup> Iwasaki and Ingkaphirom (2005: 15-16) recognize four more patterns, namely [S+ copulative /pen/ or /khue/+ predicate nominal], [S+V+DO+IO] and [T+S+V-intrans] and [T+V-trans+S]. The last two patterns are called topic-prominent structures.

Single Thai clauses can constitute single, simple sentences, or they can be concatenated to create compound sentences. Sriyapai (2013: 167-170) divides Thai sentences into three types based on sentence structure, namely simple sentences, compound sentences and complex sentences. Thai sentences, like English sentences, can have declarative, imperative or interrogative mood and positive or negative polarity, to use Halliday's terms (1994) (cf. Iwasaki and Iwasaki and Ingkaphirom (2005: 17), who conflate mood and polarity in their classification). As already indicated in our discussion of verbs, no tense or agreement markings appear in Thai clauses and time references are expressed by time adverbs or are interpreted from the context (Singhapreecha 2007:1497).

## **Transcription and Data Presentation**

Some final comments are made here about the transcription of Thai, and the presentation of data from our corpora in this thesis.

According to Kanchanawan (2006), six different systems are used to transcribe Thai words, namely the King Rama VI system (1913), the first Royal Institute system (1939), the second Royal Institute system (1968), the third Royal Institute system (1999), the Anglicization system and the ISO system (1998). The most recent ones are the third Royal Institute system (1999) and the ISO system (1998). In fact, the ISO system can be used globally as a means of communication of written messages in a form which permits the automatic transmission of written scripts either by humans and machines; however, this system still cannot always indicate the correct pronunciation of Thai words, due to the differences in the sets of written symbols between English and Thai (Kanchanawan 2006). In addition, diacritical markings are used in this system to render tones. For the purposes of the thesis, we instead employ the third Royal Institute system because it is more comprehensible for general audiences due to the lack of diacritical markings. Since we do not require an intimate understanding of Thai phonetics or

---

<sup>1</sup> In this section, the following abbreviations are used: S (subject), T (topic), V (verb), O (object), DO (direct object) and IO (indirect object), trans (transitive), and intrans (intransitive).



phonology, the use of this system suffices. The transcription table used in the third Royal Institute system, and thus this thesis, is presented in Appendix C.

In this thesis, Thai lexemes are presented with an accompanying phonetic transcription (using the third Royal Institute system, and presented between two forward slashes) and an English-language gloss (in parentheses), as follows:

Thai lexeme /phonetic transcription/ (English gloss)

e.g. ต้อง /tong/ (must/shall/have to)

In some instances where an English gloss is not applicable, Thai lexemes are followed by a phonetic transcription only.

e.g. ชุก /thuk/

In instances where the word is made of two or more free morphemes with different meanings, we add a plus sign (+) between the phonetic transcriptions of the first and the second morphemes, and so on, and a plus sign is used between English glosses, as follows:

e.g. จะต้อง /cha+/tong/ (will+ must)

The (-) hyphen is used to separate phonetic transcriptions when a lexeme consists of two or more syllables, for example:

e.g. สามารถ /sa-mart/ (can)

All examples in analysis chapters begin with the source text in English, followed by its translation into Thai and then a back translation of the Thai. The back translation is provided so that readers can follow the discussion. Back translations here follow Baker (2018:7), who translates text back into the source language from which the target language example was originally translated, as literally as possible, simply to indicate the structure of the foreign language sentence. Such back translations act like very primitive linguistic glosses, and, in our thesis, almost inevitably make for ungrammatical English.

Finally, like Baker (2018), we take a descriptive approach to examples from the source and translated texts we use in this research, and by no means seek to find fault with the authentic translations in our corpus.

# Chapter 1: Legal Language and Legal Translation

## Introduction

This chapter will mainly concern legal language and legal translation. Since legal translation is bound up in the conventions and standards of legal language, the first part of this chapter will be an account of legal language. This is followed by an overview of legal translation.

In the first part, definitions and typologies of legal language in the broad sense are presented. The treaty, a specific type of written legal text that will be investigated in this thesis, is then described, as are the characteristics of legal English and legal Thai. In the second part, typologies and characteristics of legal translation based on those of legal language will be explained since they are closely related to each other. The two main characteristics of legal texts, namely their system-bound nature and linguistic peculiarities, are presented as the main difficulties when translating legal texts. Four factors which, it is claimed (Šarčević 1997), legal translators have to take into consideration when translating law are then considered. These are: equivalence, fidelity, creativity and interpretation. In the concluding remarks, we discuss the limitations we have found in the literature and the areas to which this thesis could contribute.

## Legal Language

Prior to discussing legal translation, it is worth looking at the underlying structure of legal language, which is the main factor contributing to the former's distinctiveness. According to Coulthard and Johnson (2007: 8), language used in legal settings has been discussed and defined in a range of scholarly works. The first book mentioned is *The Language of the Law*, by law professor David Mellinkoff and first published in 1963. Mellinkoff (2004: 3) defines the language of the law as "the customary language used by lawyers in common law jurisdictions where English is the official language". Mellinkoff's (2004) book covers the historical development of legal English, and some of the grammatical features of this style of language as well as its social and cultural significance (Schane 2008: 5). Nevertheless, Mellinkoff's treatment is limited by the fact that his emphasis is on the English language only, and English common law, which is one of his areas of expertise.

Several decades after Mellinkoff's definition had been established, linguists turned their attention to language in legal settings (ibid.). Biel and Engberg (2013) claim that the study of language usage in the sphere of law has enjoyed a high degree of interest in recent years. Nevertheless, the term for such language remains problematic. Two terms which are commonly employed when talking about language in legal settings are 'the language of law' and 'legal language'.

Some scholars attempt to make a distinction between the two. For example, Burukina (2012: 712) claims that 'legal language' refers to language used when talking about legal matters, while 'the language of the law' is used with written legal documents. However, recently, these two terms have been widely used interchangeably in a generic sense to cover all types of legal language (Tiersma 1999: 139).

Legal language continues to be defined by individual authors both in the broad and the narrow sense. Cao (2007: 9) expands the scope of legal language by defining it as the language of law, language about law, and the language used in the legal process. Based on this definition, Cao (ibid.) asserts that legal language covers language used in any legal communicative situation. While Cao makes a distinction between three aspects of legal language, Coulthard and Johnson (2007: 8) pay attention to only two of these: the language of the law and the language used in legal communication. Šarčević (1997), however, considers legal language as special-purpose communication among specialists only, thereby excluding communication between lawyers and laypersons.

For the purposes of this thesis, we follow Tiersma's (1999) usage, and use 'legal language' to refer to all types of language in legal settings.

### **Typologies of Legal Language**

Legal language has fuzzy rather than discrete boundaries, with high internal differentiation (Biel 2014: 19). It is to some degree system-dependent, and classifications of legal language are often motivated by research needs and perspectives (ibid.). Likewise, Mattila (2006:4) argues that classifications of legal language are relative. In a broad sense, legal language can be divided into two major areas: monologic, which is the written language of legislation; and interactive/dynamic which is spoken in the legal process, particularly in the courtroom (Gibbons 2003: 15). Since the main concern of this thesis is legal translation, and translations of treaties in

particular, we focus first below on written legal texts and their classification; we will then home in on the particular characteristics of treaties.

Written legal texts can be categorized in either simple or more elaborate classifications according to many criteria. Some scholars categorize texts based on their function, for example, Tiersma's (1999: 139) division into: (i) operative documents, which establish the legal framework and include legislation, pleadings, petitions, judgments and private documents such as contracts and wills; (ii) expository documents, which explain the law, usually objectively, and include office memoranda and educational material about the law; and (iii) persuasive documents, such as submissions designed to convince a court. Alternatively, Šarčević (1997:11) classifies legal texts as: (i) primarily prescriptive texts such as statutes, contracts, treaties; (ii) primarily descriptive but also prescriptive texts such as judicial decisions and pleadings; and (iii) purely descriptive texts such as law textbooks. Some scholars classify texts according to the context of production. For instance, Cao's (2010: 79) four categories are: (i) legislative texts such as statutes, laws, and treaties; (ii) texts produced by judicial officers; (iii) legal texts produced by scholars; and (iv) private legal texts such as private agreements, leases and wills.

While these classifications establish categories based on the properties of the texts largely without regard to translation, Garzone (2000) discusses four types of legal written texts based on different pragmatic criteria and considering the legal force of the original and of the translated texts respectively. These are: (i) texts that shape the framework of a single national legal system where translated texts are not authoritative; (ii) documents written in a bi-lingual and/or bi-juridical country, Canadian legislation for example, where source texts and translated texts are considered official authentic texts; (iii) hybrid texts, i.e., texts produced in a supranational multicultural discourse community where all translations in the official languages are equally authentic and presumed to have the same meaning, EU legislation for instance; and (iv) international documents between private subjects in different nations where an original text agreed between parties is not necessarily authoritative and, in this case, an international document can be interpreted under the law governing it.

## **Treaties**

In this thesis, international treaties in English and their Thai translations are investigated; therefore, it is worth looking at definitions and the basic structure of treaties, which are relevant to the texts under investigation.

Generally speaking, a treaty is a generic term for all types of international agreements (Šarčević 1997: 131). In the very formal sense, Article 2(1) (a) of the Vienna Convention on the Law of Treaties, 1969, specifies that a treaty is “an international agreement concluded between States in written form and governed by international law, whether embodied in a single instrument or in two or more related instruments whatever its particular designation.”

Referring back to the classifications of legal texts proposed by scholars in the previous sections, a treaty can be categorized as one of Tiersma’s (1999) operative texts, Šarčević’s (1997) prescriptive texts and Cao’s (2010) legislative texts. Furthermore, considering Garzone’s (2000) criterion regarding the legal force of original texts and translations, only the English language versions of the treaties under investigation here are considered authoritative texts, as the Thai translations have no legal force.

In this thesis, the full texts of the treaties are analysed; therefore, a brief description of a treaty’s components is helpful. (The reasons why full texts are employed are further elaborated upon in the Methodology Chapter.)

With regard to the structure of treaties, Aust (2003: 332) suggests that most treaties consist of a single main instrument which follows a well-established pattern: title, preamble, main text, final clauses, testimonium and signature blocks and annexes (if any), (see also Šarčević (1997: 131-133) who does not, however, mention annexes). The explanation of functions and characteristics of each section of the treaty here will be based on Aust (2003) and Šarčević (1997).

Aust (2003: 332) states that the title consists of two elements: designation or name given to the treaty and a description of its purpose. There is no consistent practice in the naming of treaties; ‘agreement’, ‘convention’, ‘treaty’, ‘charter’, ‘covenant’ are the most common names used (ibid.: 333). With respect to the description of the purpose of a treaty, Aust (ibid.) claims that it should be very simple. The treaty should be known

by a shorthand title unless there is some important political reason for a long title (ibid.: 334).

According to Šarčević (1997: 131), the title serves as a means of identification while the preamble specifies the purpose of a treaty. The entire preamble is formulated as a single sentence (ibid.: 131-132). The preamble often begins with 'The State Parties to this Agreement/ Convention/ Treaty...'. Each preambular paragraph begins with a participle such as 'Recalling', 'Recognising', 'Noting', 'Convinced', and suchlike; in addition, the preamble, regardless of its length, should end with a final agreement such as, 'Have agreed as follows:' (Aust 2003: 337-338). Although preambles have their uses, their value is insignificant compared with the rest of the treaty since their primary aim is to introduce the main text by introducing the background and purpose of the treaty, so preambles should be as short as possible and not inconsistent with the main text (ibid.).

The heart of the treaty is the main text (ibid.). The main body of a treaty contains legal rules setting forth the legal prohibitions and rights of the parties (Šarčević 1997: 132). Ideally, the articles in the main text should be kept as short as possible since treaties are tools for international coordination and are in constant use, so they should be user-friendly (Aust 2003: 338-339). Nevertheless, sometimes treaties which have been drafted with short and simple articles are added to many times until the treaties become lengthy and complex (ibid.).

Regarding the final clauses, Šarčević (1997: 132), on the one hand, suggests that they consist mainly of procedural formalities and are always expressed in standard formulae. Aust (2003: 345), on the other hand, argues that, although final clauses are usually perceived as purely formal provisions copied or adapted from another similar treaty, they play an essential role since some clauses still need policy decisions.

The last section of the treaty mentioned by Aust (2003) is annexes. According to Aust (2003: 355), if there is an attachment to the treaty and it is an integral part, then it should be called 'annex' (ibid.).

In fact, a treaty does not have to be in any particular form since there are no rigid formulae (Aust 2003: 332); however, the structure of the treaty specified above is, to our knowledge, always employed when a treaty is written. The treaties under

investigation in this thesis have the same standard structure and consist of title, preamble, main text and final clauses; some treaties have annexes, while some do not.

## **Legal English**

In this section, the main features of legal English that are often perceived as a source of difficulty are first explained. This is followed by a discussion of the status of legal English in the global context.

### ***Main Features of Legal English***

Legal English is a complex type of discourse (Alcaraz and Hughes 2002: 4-14). Mellinkoff (2004: 11) lists some of the main features of legal English, namely: frequent use of common words with uncommon meanings; frequent use of Old English, Middle English, and Latin words; frequent use of terms of art, argot and formal words; and deliberate use of words and expressions with flexible meanings.

Alcaraz and Hughes (2002: 16-22) also realize the eccentric nature of legal vocabulary. They (ibid.) classify legal vocabulary into two types: symbolic items and functional items. Symbolic items include all terms that refer to things or ideas found in the real world, be they physical or mental, and which can be subdivided into three subgroups: (i) purely technical terms, (ii) semi-technical or mixed terms, and (iii) everyday vocabulary (ibid.). They claim that the second group of terms is the most difficult to recognize since they consist of common words which have acquired additional meanings in the context of legal activity (ibid.).

Functional items, on the other hand, are grammatical words or phrases that have no direct referents either in physical reality or in the universe of concepts, but which serve to bind together and order symbolic items. Functional items include ‘subject to’, ‘hereinafter’, ‘whereas’, etc. Deictics, articles, auxiliaries, modals and other purely syntactic and morphological markers also belong to this group, as do more complex units like ‘unless otherwise stated’, ‘as in section 2 above’, ‘in accordance with order 14’ and similar phrases (ibid.).

At the syntactic level, Alcaraz and Hughes (2002:18-22) specify that legal texts always contain unusually long sentences and an abundance of restrictive connectors (e.g. ‘notwithstanding’, ‘pursuant to’, ‘whereas’), abundant use of passive voice to suppress



the identity of the agent responsible for the performance of the act and abundant use of conditionals and hypothetical formulations.

### ***Use of Legal English***

For most scholars working on legal language, the main domain of analysis is legal English. As a law professor in the United States, it may not be peculiar for Mellinkoff to pay exclusive attention to legal English, but legal English is used well beyond Anglophone borders, and it is evident that legal English plays a vital role in a world legal context. Indeed, over the last decades, there has been growing concern (Alcaraz and Hughes 2002) about the use of legal English in non-English speaking countries. In African countries, despite policies promoting the use of African languages in the realm of law, English is still an important working language (Ajulo 1997: 27). In Malaysia, English is frequently used in the High Court and dominates the Court of Appeals and Federal Courts; it is also the default language of most commercial law and much civil litigation (Powell and Hashim 2011). The question thus arises as to why and how legal English has become so important in the global jurisdiction. The first possible answer can be that the use of English is an inheritance of colonization in some African countries and Malaysia.

The prevalence of English in the legal world is also influenced by globalization. According to Alcaraz and Hughes (2002: 2), given the prominence of English as the language of contemporary communication and trade, there has been an increase in the use of English in all sectors, including special-purpose communication such as international trade and law. Furthermore, due to the need for a common language within an increasingly disrupted and globalized world, English has become a cosmopolitan language, with which lawyers feel at ease although it is not their mother tongue (Nadrag et al 2013: 846).

In addition, the use of legal English is also encouraged in many countries without a history of colonization, Thailand, for instance. In Thailand, English is not officially used in legal communicative situations in general; however it has played an important role in such situations for several hundred years.

### **Legal Thai**

This section aims to provide an overview of Thai legal language. A brief introduction to the history and development of Thai law is first given to provide a basic

understanding of how the legal system of Thailand has been shaped and of how the English language has influenced Thai legal language to the present day. The main features of legal Thai are then illustrated.

It is necessary to clarify that research on legal language and legal translation in Thailand is very rare. The existing literature is very limited and covers just a small part of the whole picture. According to Boonchalermwipas (2009: 5), although the study of Thai legal history has long been established, there are very few documents in this area, the majority of which were written by foreigners and some of which are stored at the National Archives of Thailand.

Furthermore, Thai legal language is even more under-researched than Thai legal history. To our knowledge, there is only one book on the subject, namely, *Thai Legal Language*, written by Thanin Kraivixien (2012), a former Prime Minister of Thailand.

### ***Thai Legal Periods***

It should be noted that Thailand was known as ‘Siam’ until 1989, but for the sake of simplicity, we use the term ‘Thailand’ to refer to both Siam and Thailand.

Thailand’s history can be divided into four periods based on the ruling dynasties, namely the (i) Sukhothai (1238-1376), (ii) Ayutthaya (1350-1767), (iii) Thonburi (1767-1782) and (iv) Rattanakosin or Bangkok (1782- present) periods. In each period Thailand has been ruled by a different dynasty and established a different capital city. For example, the Rattanakosin or Bangkok period began in the reign of King Rama I, the founder of the Chakri dynasty, while the current King of Thailand is King Rama X from the same dynasty and the capital city is Bangkok. Traditionally, Thai legal history is classified according to this period division because it seems convenient to categorize Thai legal history based on its national historiography (Boonchalermwipas 2009: 47).

Nevertheless, this categorization of Thai legal history has not always been recognized as the most appropriate, since the characteristics of the law enforced in each period have not been the main focus. Boonchalermwipas (2009: 48) suggests that Thai legal history should be divided in line with the types of law enforced irrespective of dynastic periods. He (ibid.: 49-51) proposes two periods of Thai legal history, based on the most significant change to the legal system, when the law was reformed and modernized. His periods are thus: the pre-modern-law period (the period before the reign of King Rama

IV) and the modern-law period (the period from the reign of King Rama IV onwards). The reign of King Rama IV (1851-1868) is considered as a new period because it was when Thailand started receiving and incorporating concepts from Western law (Kasetsiri and Pongpanitanont 2005: 55).

The two laws enforced in the pre-modern-law period were indigenous law (which had been used since the Sukhothai period) and Indian law (which was adopted later) (Boonchalermwipas 2009: 50-52). In fact, Thailand established contact with the West in trade and official matters in the reign of King Rama III (1824-1851); nevertheless, westerners refused to follow Thai law claiming that it lagged behind their modern world and that both trials and punishments were cruel and barbaric (ibid.: 140).

During the European colonial era, King Rama IV was aware that this uncivilized legal system could have been used by potential colonizers as a reason to subjugate Thailand; among other compromises, the King thus appeased potential colonizing countries by committing to treaties with them (Kasetsiri and Pongpanitanont 2005: 58). Pursuant to the treaties, apart from trading freely in Thailand and paying very low taxes, foreigners whose countries had entered into these treaties were entitled to extraterritorial jurisdiction (ibid.). All of these concessions put Thailand at a disadvantage (ibid.).

### ***Foundation of Modern Thai Law***

As part of his attempt to preserve independence and to reform Thai indigenous law, King Rama IV restructured and modernized the Thai legal system by incorporating Western legal concepts into the country's legal system (Boonchalermwipas 2009: 137-140). In the beginning, some parts of English, French and German law were incorporated (Kraivixien 2012:44-65).

Then, in the reign of King Rama V (1868-1910), there was a significant reformation in the Thai legal system. King Rama V repealed many parts of indigenous law and decided to draft a new legal framework based on Western law (ibid.: 165). In the drafting process, the King formed a drafting committee by hiring many foreign law experts to work with Thai experts. The King and the committee decided to use (European) Continental law or code law as a prototype (ibid.).

Regarding the codification, the English language played a vital role. The English language was used in legal drafting because it was more convenient for the foreign and

Thai drafters since they had a good understanding of English (ibid.: 44). However, according to Kraivixien (2012: 45), the legal drafting committee did not use legal English, realizing its complicated characteristics; rather, the type of English language adopted in legal drafting was similar to the Queen's English but it was more concise and precise; as such it is considered as 'Functional English' or 'Technical English'.

After the law was drafted in English, it was translated into Thai. There were also many obstacles which translators had to face while translating the English draft into Thai; for example, translators had to invent many new words since there were many English words for which translators could not find equivalents in Thai (ibid.: 166). On some occasions, transliteration was employed (ibid.: 169). In addition, the translators had to avoid or minimize the inappropriate influence of the English language on Thai law; for example, passive voice which was prolifically used in English legal texts seemed to be barred in Thai versions because the passive voice in Thai was mostly used to convey negative meanings (ibid.: 171).

Although the English draft was translated carefully, there were some cases where translators mistranslated, resulting in changes in content, such as those in Sections 442 and 218 of the Civil and Commercial Code, which are still enforced today (ibid.: 168). This can be problematic because only the Thai texts are considered authoritative.

### ***Features of Modern Legal Thai***

Today, Thai law is drafted in the Thai language, but the influence of the English language over Thai written legal texts still remains. Kraivixien (2012: 43-50) points out that the desirable effects of Functional English used in the first drafts of Thai legal texts continue to the present day. Kraivixien (ibid.) explains that since Thai legal texts were first translated from the Functional English version which was concise, clear and easy to understand, the Thai version was also concise and clear which was unlike the Thai legal language used in former indigenous law and the Thai language in general. Furthermore, this type of modern legal Thai continues to serve as a model for legal Thai to the present day (ibid.).

At the lexical level, Kraivixien (2012: 173-177) divides Thai legal vocabulary into three types: purely technical terms, terms of art, and idiomatic expressions. Purely technical terms are the terms found and used only in the legal context (ibid.). Terms of art or 'words of art' refer to words or phrases that have a precise or specialized meaning

within a particular area. In general, idiomatic expressions are expressions which have a different combined meaning from the meanings of the individual words in the expression, and legal idiomatic expressions are idiomatic expressions used in the legal context (Kraivixien 2012: 175-176). Legal idiomatic expressions can cause confusion and misinterpretation for both lay persons and legal experts since sometimes it is difficult to determine whether the expression is being used idiomatically or literally on any given occasion and if the expression is being used idiomatically whether it is being used in its 'normal' idiomatic meaning or whether it has specialized idiomatic meaning in legal contexts (ibid.).

While the first two types of Thai legal vocabulary specified above are similar to those of English specified by Alcaraz and Hughes (2002: 16-17), namely English pure technical terms and semi-technical mixed terms, respectively, the third type of Thai legal vocabulary does not match any category of English legal vocabulary proposed by Alcaraz and Hughes (2002).

### **Legal Translation**

Based on traditional categorization (Cao 2013), legal translation is labelled as a type of specialized translation. Many scholars consider legal translation a category in its own right (Garzone 2000: 395; Harvey 2002: 177), one that draws attention from both legal and translation scholars who set out to investigate its typologies and characteristics.

#### ***Typologies of Legal Translation***

According to Biel (2014: 50-51), it is notoriously difficult to classify legal translation through traditional specialized-translation criteria, however. Fuzziness in typologies of legal texts leads to fuzziness in typologies of legal translation (Biel 2014: 50). Nevertheless, a number of scholars have tried to classify legal translation into subtypes. Since typologies of legal texts vary according to the user's perspective, such aspects are inherited by typologies of legal translation as well.

The majority of classifications of legal translation have focused on the source text; thus, to some degree they tend to reflect classifications of legal texts (Biel 2014: 51). Šarčević (1997: 11) thus classifies legal translation into three types, according to the function of the source text (see above) as follows: (i) purely prescriptive, such as translations of laws and regulations; (ii) primarily descriptive but also prescriptive, such

as translations of judicial decisions; and (iii) purely descriptive, such as translations of law textbooks.

Nevertheless, Cao (2007: 9) argues that these traditional classifications are incomplete and could pose problems for two reasons. First, these classifications rely heavily on attributes of the source such as the function or the status of the source text; thus they focus inadequately on the target language variables, e.g., status or function of the target texts. Second, some traditional classifications, for example, Šarčević's (1997) classifications, exclude some documents used in actual legal communicative situations since they are not considered to be legal texts in the first place.

Cao (2007) thus proposes a new typology of legal translation according to the purposes of the translated texts, namely, (i) normative purposes, such as the translations of equally authentic legal texts in bilingual or multilingual jurisdictions; (ii) informative purposes, e.g., translations of legal documents providing information to target text readers; and (iii) general legal purposes, e.g., the translation of general texts used in legal settings. Cao (2007) claims that this classification seems to cover translations produced in all legal communication.

Biel (2014: 51) states that according to the status of the target text, legal translation can be classified into two types: (i) authoritative translations which are equally authentic to the source text; and (ii) non-authoritative translations which are not legally binding. In some cases a translated legal text will tend to have a lower status than the original. This applies in particular to legal texts translated in monolingual countries where the status of the target text is 'only' that of a translation and the original prevails over it (Biel 2014: 52). However, in some cases, the status of such translated texts can coincide with that of the source texts, such as legal texts translated in bi- and multilingual countries where legal translations must be as authentic as the original.

In this thesis, Biel's (2014) typology applies according to which only the legal texts written in English are considered authoritative texts while the Thai translations have no legal force.

### ***Characteristics of Legal Translation as a Source of Difficulty***

Both legal language and legal translation are marked off from ordinary language and recognized as language for special purposes. As Šarčević (1997: 66) posits however,

while other special-purpose languages, like mathematics and chemistry, can be universal and largely uniform throughout the world, this is not the case in the field of law. According to Šarčević (1997), legal language is deprived of the universalism of other types of special-purpose language and such characteristics make translating legal texts more difficult than translating other special-purpose texts. While Montgomery (2000: 253-254) argues that the presumed universalism of scientific discourses such as mathematics and chemistry is a myth, Šarčević's (1997) claim that the lack of universality of legal texts is the real source of difficulty in translating the law remains valid. In the following sections, we discuss two main characteristics of legal texts that are sources of difficulty in translation and which arise from the lack of universality in the legal realm, namely the system-bound nature of the law and its linguistic peculiarities.

#### System-Bound Nature

Legal language is bound to individual legal systems. Though there are three major world systems (civil law, criminal law and religious law), there are hundreds more specific legal systems employed throughout the world. Each nation-state has its own legal system and the structures and characteristics of these systems are highly variable (Cabanellas 2014: 3). This is because each legal system has its own history, principles, and patterns of reasoning because it has been designed to answer the requirements of a certain nation and reflect its idiosyncrasies (Biel 2014: 49). Legal systems are peculiar to the societies in which they have been formed, as is legal language. Since legal language is a product of the legal system, it cannot be universal but is tied to a national legal system (Weisflog 1987: 203 in Cao 2007: 23). As a result, legal terms and concepts which are the product of the legal system have a system-bound nature (Šarčević 1997: 232). Despite its system-oriented nature, there have been many attempts to globalize and harmonize legal concepts in some areas of law (Biel 2014: 50). Even if a national legal system conforms more or less to a major legal system, such as common law or civil law, it can still have different systems of reference, for example, a term in one legal system can refer to a similar or different concept in another legal system. Therefore, translators are required to determine within which specific legal system the texts and concepts are to be interpreted (Šarčević 1997: 67), since the meaning of legal texts is comprehensible only within their social context (Sumner 1979: 277 in Šarčević 1997: 69).

## Linguistic Peculiarities

Legal language is a linguistic phenomenon in its own right, because of its peculiarities in vocabulary and sentence structure. This statement is acknowledged by both legal and translation scholars. In the following sections we outline two of the main features of legal language, namely its technicality and indeterminacy as identified in the literature.

### Technicality

Gibbons (2003: 39) asserts that every field of expertise develops its own language features, and this applies equally to the realm of law, since technicality is strongly marked in legal language. Cao (2007: 20-23) classifies peculiar linguistic features of legal language into four categories: (i) lexicon: complex and unique legal vocabulary, (ii) syntax: complex and long structures and syntactic peculiarities, and (iii) pragmatics: the performative nature of language, and (iv) style. Bhatia (1994) states that legal documents are assembled on the basis of a type of ‘cognitive structuring’ where core statements along with all details are normally presented as a single syntactic structure, often with great length and complexity.

From the legal point of view, the technicality of legal language is not always considered a disadvantage. Gibbons (2003: 36-37) suggests that there are a range of processes and institutions that are unique to the legal system, so a specialist language is required when referring to them. If the technical terms did not exist, there would be a need for long and clumsy explanations each time reference was made to the technical process or notion (*ibid.*). Nevertheless, from the translation studies’ point of view, such technicality can pose problems. Much of the technicality of legal language is a product of the need for maximal precision (Gibbons 2004: 2), but technical words can be as unclear as any other words (Solan 1993: 132).

### Indeterminacy

Indeterminacy characterizes all three disciplines involved in legal translation: linguistics, translation theory and law. Indeterminacy is an umbrella term for certain properties of natural languages, namely, ambiguity, vagueness and generality (Paunio and Lindroos-Hovinheimo 2010: 396). Generally, ambiguity and vagueness are two concepts repeatedly cited when authors mention the characteristics of legal language and translation. Terms are considered ambiguous when they have more than one meaning, while concepts are vague when they have an imprecise meaning. Cao (2007:



73) states that there are three sources of indeterminacy when translating legal texts, namely, lexical indeterminacy, syntactic ambiguity and indeterminacy arising from errors. Alcaraz and Hughes (2002: 30-45) also mention lexical vagueness and syntactic ambiguity as the main problematic characteristics when translating law.

Apart from accidental clashes between the ideal perception of legal language (as precise and univocal) and the actual indeterminate nature of legal language (Marinetti 2008:122), translators are also required to find out if the indeterminacy exists on purpose. Indeterminacy can be intentional since it is sometimes viewed as a virtue and could be retained in translation since it can be used to reach a compromise or create uncertainties which a party will seek to exploit (Harvey 2002). Though vagueness is positive when it provides a certain amount of flexibility needed to guarantee applicability of the law in a variety of situations, exact translation is made difficult by vague texts (Paunio and Lindroos-Hovinheimo 2010: 397). The problem of translating indeterminacy is complicated by translators' varying competence in understanding the subtleties involved (Harvey 2002: 181). Translators must be competent enough to distinguish intentional from unintentional indeterminacy since an attempt to clarify intentional indeterminacy might undermine the source text's intent (Balaza 2014: 360).

It is presumed that though major differences exist in different languages due to differences in legal systems, the aforementioned linguistic peculiarities are common in most if not all legal languages, and to varying degrees (Cao 2007: 23).

### **Translating Law as a Problem-Solving Process**

Accurate translations of legal texts are elusive goals which are never unambiguously achieved. Legal translation is difficult to practise because it operates between different languages and also distinct legal systems. This incongruity is considered one of the main obstacles for translators of legal texts. In addition, translators face the conflict between accuracy and naturalness in legal translation. To sum up, to produce accurate translations, legal translators face both system-specific and language-specific problems. Newmark (1982 in Šarčević 1997: 161) concludes that legal translation is more restricted than any other form, and studies of legal translation also shed light on the decision-making process of legal translators (ibid.). Translating law is a form of cultural interaction since it represents a bridge in understanding legal texts in different legal cultures, and the key individuals are translators. This challenging task is imposed on

translators who are no longer perceived as text drafters but text producers and cultural mediators (Šarčević 1997). The question is over the factors the translators can use to make their decisions to produce their work and to what extent. As in many other areas of translation, the nature of problem solving in legal translation is often cast as one of finding ‘equivalence’, assuring ‘fidelity’ or using one’s ‘creativity’, topics which are treated below.

### ***Equivalence***

The concept of equivalence in translation theory is as common as it is vague and controversial; how to achieve it in actual practice has been one of the most widely debated issues (Garzone 2000: 5). Various notions and typologies of equivalence in translation have emerged in different paradigms of translation studies. In addition, the scope and focus of equivalence has always changed over time.

Traditionally, linguistically-oriented translation scholars (e.g., Nida: 1964) interpreted translation equivalence as an approximate or even equivalent translation. However, some translation scholars contend that sometimes it is unfair to believe that translations could convey the same, stable and language-independent meaning as their source texts (Kenny 2009: 96).

Some translation scholars deploy the concept of equivalence as they deem appropriate. For instance, Toury (1995) proposed the notion of equivalence within descriptive translation studies as different from a particular target-source relationship established on the basis of a recurring type of invariant; rather, it is a functional-relational concept where equivalence can be described as any kind of relation between sources and translated texts. Likewise, various typologies of translation equivalence have emerged as well.

According to Kenny (2009: 97), equivalence can be categorized into many types following several scholarly models, for instance: (i) Koller’s referential or denotative equivalence, connotative equivalence, text-normative equivalence and pragmatic equivalence (1979 and 1989); (ii) Nida’s formal equivalence and dynamic equivalence (1964); (iii) Baker’s textual equivalence (1992); and (iv) Newman’s functional equivalence (1994). This reflects that over time there have been shifts in the notions and typologies of equivalence in translation theory.

## Legal Equivalence

The shift toward theories and typologies of equivalence also occurred in legal translation practice. Traditionally, legal translation was characterized by the unquestioned predominance of what Šarčević (1997: 24) calls ‘strict literalism’, an emphasis on the preservation of the letter rather than the effective meaning of target texts (Garzone 2008: 50).

Previously, when dealing with legal texts, many legal translators employed formal equivalence when they felt uncertain in making a linguistic decision, fearing that any linguistic changes might distort the content; therefore, a translator should avoid any possible risks (Šarčević 1997: 116). Today, according to Weston (1991: 25 in Biel 2014: 43), a formal equivalent can be useful and is used in legal translation when the meaning is transparent. In fact, this technique is seen frequently in EU translations when a signal of the supranational nature of a term is required (ibid.).

A criterion of equivalence specific to legal translation has also emerged. Beaupré (1986: 179) proposes ‘legal equivalence’, in consideration of the legal effects that a translated text will have in the target culture. Sager (1993: 180) explains that the concept of legal equivalence will seek to achieve both identity of meaning and identity of the legal effects of the source text. Basically, the criterion of legal equivalence is similar to that of Nida’s (1964) dynamic equivalence and pragmatic equivalence (Garzone 2000: 5).

Nevertheless, this new notion of equivalence poses a new crisis: whether the concept of legal equivalence can be applied to all actual legal translation practices or not. Garzone (2000) investigates how much legal equivalence can be applied to translated legal texts generated in three different legal settings, namely the legal texts produced in single, bi-, and multi-juridical countries. Garzone (ibid.) discovers that while legal equivalence can be used when translating legal texts in bi- and multi-juridical countries, the approach adopted when translating texts produced in a single national system is usually formal since the translated text has no legal validity of its own; therefore, glosses and commentaries can be used to indicate the original meaning. Garzone (2008: 51) concludes that legal equivalence is not universally applicable, which means that under certain circumstances the strict literal approach is suitable when translating certain legal texts.

## Equal Intent

In addition to the traditional equivalences mentioned above, Šarčević also proposes another type of equivalence. Because of the special characteristics of legal language, Šarčević (1997: 70) recognizes that the main goal of legal translations cannot be the same as that of other types of technical translation. Legal translators cannot produce translated texts with the same meaning as the source text. First, she defines a goal that legal translators should have: producing a text with the same legal effect. Later, she proposes a redefined goal of legal translation which must contain the same intent as the source text. Šarčević (1997: 73) claims that while equal effect has priority over equal meaning, both are inferior to equal intent.

According to Šarčević (1997: 73), there are two forms of equal intent: macro and micro intents. The macro-intent of a text is identified with its general communicative function; for example, legal instruments such as laws and contracts are informative to some extent but their primary function is imperative (ibid.: 9-10). The micro-intent is the underlying purpose of the particular text, namely what the author is specifically attempting to achieve (ibid.: 73). While it is believed that special-purpose texts consist purely of facts, the meaning of which is not influenced by author intent, this is not true of legal texts; for instance in treaties between states, the micro-intent is known as the intent of the States parties (ibid.). Šarčević (1997: 73) claims that this intent is the true and original intent of the texts. In this sense, Šarčević (1997) believes that legal translators must strive to produce a text which conveys both macro- and micro-intent.

The micro-intent is referred to as the author intent by Neubert and Shreve (1992: 72) who states that all texts are written to achieve some result intended by authors and translations should convey such intention. Katan (2014: 172) argues that although the idea that one needs to understand the underlying intention of a writer to translate effectively has become a corner stone of translation theory, it has been criticized for many reasons, including the fact that many writers are not entirely conscious about their intentions. Likewise, Gentzler (2001: 57) contends that there is a problem arising from the devotion to and dependence upon the original author's intention. The problem is what literary critics referred to as the 'intentional fallacy', which means what a work says and what the author intended to say are two different things (ibid.). Nevertheless, it should be noted that an argument over 'intentional fallacy' is relating to only literary works; to our knowledge, it is not mentioned in literature on legal translation.

## Fidelity

The crisis between fidelity and transparency has been a central debate in translation studies. This conflict is widespread in all types of translation, but is strongly marked and prominent in legal translation (Biel 2014: 49). Traditionally, fidelity in translation referred to scrupulous adherence to source texts; however, recently, the concept of fidelity has been seen as relative to more than the source text alone and has been largely redefined from individual scholars' perspectives (Harvey 2002: 180).

In the realm of legal translation, Šarčević regards fidelity as the 'factual' accuracy of translations compared to source texts and holds that accuracy is of utmost importance in legal translation (1997: 670). In other words, substance must always prevail over form in legal translation (Šarčević 2000: 5). Vermeer and his advocates, on the other hand, relate translation fidelity not to the meaning of the source text but to the function of the target text (Garzone 2000). In fact, Vermeer attempted to apply the functional approach to language for special purpose translation including legal translation, claiming that it is possible to apply this approach in translating legal texts (Vermeer 1986: 38 in Šarčević 1997: 66). Nevertheless, the application of this theory in legal translation has remained in serious doubt in the eyes of specialized translation scholars. Šarčević (1997: 65) argues that to preserve the information content is the legal translator's main goal; nevertheless, according to Skopos theory, the most important element in translation is function not content. Šarčević believes that the functional approach should be avoided in legal translation (*ibid.*). Between the objection to and support for the adoption of functionalism in the field of legal translation, some scholars find a middle ground. Garzone (2000: 9) states that some of the most controversial corollaries of functionalism have to be modified or even discarded when applying a functional approach in legal translation due to the particular pragmatic status i.e. legal force and validity. However, according to Garzone (*ibid.*) on the whole, the functional approach can be suitable for legal translation practice in certain circumstances.

Šarčević (1997) strongly objects to the functional approach because her main investigation is on translated legal texts produced in a bi-juridical country, Canada where translated texts have an authoritative status. Under such circumstances, the functional approach may be inappropriate since it might affect the legal validity of translated texts. Nevertheless, according to Garzone (2000: 5), and as already indicated, there are some types of legal texts such as international documents between private

subjects in different nations where an original text agreed between parties is not necessarily authoritative and in this case, an international document can be interpreted under the law governing it and the functional approach can be applied.

### Creativity

At first glance, creativity seems to be forbidden in the realm of legal translation where accuracy is given priority and aesthetic effects are of least importance. However, lately, creativity has been the subject of much debate in legal translation studies. The first question that arises is whether translators can be creative when translating legal texts. Pommer (2008) believes that translation is a creative activity and this property is inherent to legal translation as well. In the field of psychology, creativity is defined as the ability to produce work that is novel, appropriate, and accepted by society as being creative (Sternberg 1999: 502 cited in Pommer 2008:356). According to the aforementioned definition, translating seems not be a creative task due to the lack of novelty. However, Pommer (2008: 359) argues that many researchers have emphasized that the relationship between creativity and problem solving is very close. Translating law involves a lot of problem solving; thus, legal translators also use this skill when translating texts. Apart from Pommer's support, many scholars, e.g. Durieux (1991), Neubert (1997) and Šarčević (1997), also believe in the value of creativity and encourage it in legal translation. If legal translation is presumed to be a creative endeavour, then the next question is to what extent translators can be creative in legal texts. To answer such questions, the limitations and constraints on creativity in legal translation must be taken into account. Durieux (1991 cited in Pommer 2008: 360) contends that translators can and should be creative in technical texts, but should be more cautious when dealing with legal texts, realizing that they are subject to special restrictions. The major restrictions which all legal translators respect are the asymmetry of legal systems and legal cultural diversity resulting in incongruity in legal terms and concepts (Pommer 2008: 362). Another constraint to creativity in legal translation proposed by Šarčević (1997: 116-118) is the mandatory use of standard formulae. The use of standard formats minimizes ambiguity and facilitates the application of parallel texts where entire clauses have to be in a standardized form; hence, in this kind of text, translators are not permitted to be creative. However, in non-standardized texts, translators are allowed to be creative (ibid.).

## Interpretation

From a lawyer's standpoint, Gibbons (2003: 232) explains that the term 'interpretation' in legal contexts is reserved for the process of deriving meanings from operational documents; in other words, interpretation is solely practised by lawyers. In this regard, Pigeon (1978: 38 in Šarčević 1997: 91) affirms that lawyers agree that legal translators must understand the source text so as to produce adequate translations; nevertheless, they are not allowed to interpret the source text as lawyers do. However, translation scholars have a different perspective: Gémár (1995: 167 in Šarčević 1997: 91), for example, affirms that one of the main tasks of legal translators is interpreting the law. While Šarčević (1997: 92) agrees with the lawyers' view but contends that the translator's duty is to express what is said in the source text and not what s/he thinks it should have said, she (ibid.) also contends that it is difficult to make clear-cut boundaries between understanding and interpretation. The interpretation issue always co-occurs with ambiguity. It is generally emphasized that translators must avoid resolving ambiguity (Harvey 2002: 182). Šarčević (1997: 92) states that translators have no authority to resolve ambiguity. She raises a case of treaties where ambiguity is retained for obtaining consensus; thus, an attempt to clarify such ambiguity by the translator is one of the biggest apprehensions of treaty architects. Thus, it can be concluded that legal translators must be aware of this restriction though it is hard to distinguish between expression and interpretation.

## Conclusions

Although translations of legal texts have been far less investigated than those of literary texts, recently interest in the study of legal language and legal translation has increased. Nevertheless, literature on legal language and legal translation is still sparse. The literature on legal language and translation is also scattered and fragmented since it comes from different strands; the legal and the linguistic. Furthermore, with notable exceptions such as Biel (2014), there has been little interaction between these two disciplines. In this sense, to obtain a multi-dimensional view of legal language and legal translation, studies from as many as angles as possible are required.

So far, there have been studies of legal translation based on corpora in limited language pairs, particularly English into European languages. Nevertheless, to date no empirical research on legal language and translation has been conducted in the English-to-Thai language pair (or vice versa). Nor have many of the concepts presented in this chapter

been ‘put to the test’ in studies of translation into Asian languages in general. And much of what has been said is based on prescription rather than description. Studies such as the present one thus stand to make a novel, data-oriented contribution to studies of legal translation.



## **Chapter 2: Corpus-Based Descriptive Translation Studies and Translation Features**

### **Introduction**

Since the 1970s, translations have been considered as a unique text type; due to a new direction taken in translation studies, Descriptive Translation Studies, the idea of searching for general laws and regularities in translation has also gained much attention (Chesterman 2010: 40). To date, two views on the distinct features of translations, i.e. features which set them apart from original texts, have been presented: (i) translated texts can reflect artifacts of the translation process and (ii) translated texts exhibit traces of the original language from which the texts were translated (Lembersky 2013: 1-2).

With respect to the first view, one of the best-known advocates for seeking such regularities is Mona Baker. An explanation of the distinctive nature of translations was proposed by Baker in 1993 and 1996 in which she claimed that there are universal features of translations which can be seen as a product of constraints inherent in the translation process itself (1993: 176). Baker (1996) also provides candidates for translation universals as follows: (1) explicitation, translations tend to make explicit what is implicit in the source text; (2) simplification, translations tend to be more simplified than non-translations; (3) normalization, translations tend to exaggerate typical features of target language; and (4) levelling out, translations tend to be more alike than non-translations.

In addition to the translation features proposed above, to date there are other possible candidates that could fit in this category proposed by translation scholars from time to time. One of the interesting concepts is the unique items hypothesis proposed by Tirkkonen-Condit (2002). According to Tirkkonen-Condit (*ibid.*), unique items are target-language linguistic items without linguistic counterparts in the source language and such items are deemed to be underrepresented in translations compared to the non-translated texts in the same language.

Toury (1995) proposed two laws of translations, namely the law of interference and the law of standardization. In brief, the former law claims that translations are likely to manifest interference from the source text; the latter states that translations are likely to demonstrate a more standardized style than the original. According to Toury (1995:

275), interference is a phenomenon where the make-up of the source text tends to be transferred to the target text and this transfer can either be positive (a greater likelihood of choosing features which exist in the target language and are used in any case) or negative (deviations from normal codified target language patterns). The main contribution of Toury's work is the observation that interference is manifested not only as easily noticeable ungrammatical distortions of target language patterns, but also as more subtle distortions of the frequency of target language patterns, which are well-suited to measurement with corpus tools (Biel 2014: 107).

The concept where a source text leave traces in its translation is supported by Teich (2003). Teich (2003: 145) hypothesizes that a translation into a given language may be oriented towards the source language; that is, the source language shines through in translations, and that a translation tends to use more features of the target language; that is, normalization. Based on her corpus-based studies, she confirms that some features in English translations from German and vice versa are affected by the source language shining through and normalization (ibid.: 207).

Baker's hypothesis is somewhat similar to Toury's laws; however, Baker's notion of translation universals is much more radical in claiming that translation universals are independent of language pairs, genres, and cultures and are not a product of translation norms (Biel 2014: 97). Although Baker's hypothesis on universals seems absolute, the particular contribution of her work is that it was the first time that hypotheses about these translation properties were explicitly formulated and tests for them proposed (Teich 2003: 22), specifically after Baker's strong support in 1996 for employing corpora to investigate universals in translation.

According to Biel (2014: 95-96), corpus-based translation studies may be regarded as a paradigm shift due to the revival of linguistic methods, a shift towards empirical quantitative methods and an orientation towards target texts; in addition they have been facilitated by a shift from the relation between the source text and target text to translations as independent texts in their own right, and which function in their receiving culture. In addition, large-scale corpus-based studies have been conducted to find evidence for Baker's universals (for example, Laviosa 1998; Olohan and Baker 2000; Teich 2003; Hansen-Schirra et al. 2007; and Becher 2011).

Nevertheless, no consensus regarding the manifestation of translation features has been reached, and the issue of translation universals remains highly controversial and criticized and not all authors use the term ‘universal’. Like Toury, Bernardini and Zanettin (2004: 52-53) prefer the notion of law to that of universal, explaining that laws may be proposed that describe widely- and even universally- followed norms; unlike universals, laws in the social sciences are subject to conditioning factors of various kinds and as such would appear to be much more in tune with neo-Firthian linguistics and be much more amenable to discovery by means of corpus analyses. Currently most scholars (such as Chesterman 2010: 44; Olohan 2002: 423) prefer less strong words such as ‘typical features’, ‘tendencies’ or ‘regularities’ of translation. Therefore, for the purpose of the thesis, the term ‘translation features’ is used to refer to what some scholars have called ‘translation universals’.

In the next section, the descriptions of the three certain translation features used in this investigation, namely the notion of explicitation, unique items and source-language shining through are presented. The concept and studies of explicitation are first described, followed by those of source-language shining through and the unique items hypothesis. It should be noted that among studies dealing with translation universals, the general preference has been to focus on explicitation for a long period of time and this is why it always comes into the picture and is a prime candidate for testing translation universals (Saldanha 2008: 20). Likewise, Chesterman claims (2010: 41) that explicitation is one of the most widely studied and debated potential universals. Due to the substantial contributions of studies on explicitation, it is not unusual that the review of such a feature is longer than that of the other two features.

## **Explicitation**

This section reviews the most frequently cited definitions of explicitation in translation and its typologies and their criticisms, followed by previous studies based on the definitions and typologies (along with criticisms made of these studies), with a special focus on explicitation in Asian and African contexts. Then it also addresses different approaches to explicitation.

### ***Definition and Criticism***

Historically, the term ‘explicitation’ in translation was first defined by Vinay and Darbelnet (1958/1995: 342) as a translation technique which makes explicit in the target

text what remains implicit in the source text. Later, in 1986, Blum-Kulka investigated shifts of cohesion and coherence in translations based on contrastive analyses of a small number of source texts and translations and found that the level of cohesion in translations was higher than in the source texts. She then put forward the 'Explicitation Hypothesis' suggesting that explicitation might be a universal strategy inherent in the language mediation process (1986).

However, there is wide criticism of these definitions. According to House (2008: 11), terms like 'explicitation' or 'explicitness' are too general, and they should not be used unless one is perfectly clear about whether they can be precisely defined and operationalized. Becher (2010b: 17) criticizes the definition proposed by Vinay and Darbelnet as being extremely vague since it cannot provide a concrete explanation of what the term 'explicit' means. Becher (ibid.) realizes that it is undeniable that the first definition coined must somehow contain vagueness; nevertheless, he claims that many scholars are likely to embrace this definition without noting its vagueness.

While the original definition by Vinay and Darbelnet has been debated for its vagueness, Blum Kulka's Explicitation Hypothesis has been criticized for its overgeneralization. According to Pym (2005: 31), this hypothesis is worth close attention. Since part of the original claim is based on cohesion markers, the hypothesis formulated does not strictly concern all those uses of language that refer to things beyond the text or conversation. He explains that those cohesion markers could have no meaning beyond their function as textual pointers, and, as such, the Explicitation Hypothesis does not involve anything beyond the text and the only result of explicitation is redundancy, not explicitation. He claims that Blum-Kulka's study observes the repetition of pointers rather than a strict process of the implicit becoming explicit (ibid.). On the same grounds, Pym contends that other studies claiming to offer evidence of explicitation based on Blum-Kulka's hypothesis concern pointers of one kind or another but not explicitation. It could be noted that Pym's argument is based on linguistic metaphor.

While Pym criticized the Explicitation Hypothesis in terms of the limitations of pointers which are insufficient for claiming the existence of explicitation, Becher (2010a: 1) points out that the Explicitation Hypothesis does not even qualify as a scientific hypothesis. Becher (2010a: 6-8) raises three serious problems in connection to Blum-

Kulka's Explicitation Hypothesis by claiming that this hypothesis should be abandoned since it is "unmotivated, unparsimonious and vaguely formulated". In addition, Chesterman (2010: 41) argues that the concept of explicitation itself has proven problematic since it has been interpreted in many conflicting ways, which makes it impossible to compare results of different studies.

Thus, the major problem with the aforementioned definitions of explicitation is that these definitions are conceptualized in a way that is far from explicit. They contain a mixture of vagueness, confusion and overgeneralization. The same situation also occurs with typologies of explicitation.

### ***Typologies of Explicitation***

With regard to typologies of explicitation, the best-known classification has been proposed by Klaudy (2009: 106-7) who identified four different types of explicitation in translation as follows: (1) obligatory explicitation required by grammatical (syntactic and semantic) differences between source and target languages; (2) optional explicitation dictated by differences in text-building strategies and stylistic preferences between languages; (3) pragmatic explicitation occurring where culture-specific knowledge is presupposed in the source text; and (4) translation-inherent explicitation attributed to the nature of translation itself. It should be noted that Klaudy (1996: 103), realising that this classification must be exercised with caution, proposes to separate obligatory, optional and pragmatic explicitation, which are not universal, from translation-inherent explicitation. In other words, Klaudy (1996:33) reminds readers that a large scale contrastive stylistic study is first required to identify and differentiate optional explicitation from universal ones.

Though this classification of explicitation has been cited widely in explicitation studies, some scholars are very sceptical about this typology. According to Englund Dimitrova (2005: 38), Klaudy's account is an interesting attempt to unite the manifold observations on explicitation into a single framework; however, being built on partly different criteria (hypothetical causes versus linguistic realization), its application becomes difficult. Englund Dimitrova (ibid.) suggests that obligatory and optional explicitations are related to linguistic systems; however, the borderline between what is optional and what is obligatory can be fluid, and importantly, even obligatory options will sometimes not be realized. Furthermore, Becher (2010a: 3) notices that Klaudy

(2009) cannot present any examples for type (4) explicitations and argues that while obligatory and pragmatic explicitations can be easily identified, it is hard to distinguish between optional and translation-inherent explicitation. Biel (2014: 100) argues that this classification lacks clarity as the categories seem to be overlapping. In addition, it is hard to identify whether translation-inherent explicitation is a conscious or subconscious phenomenon (Becher 2010b: 12).

Again, the typologies illustrated above lack clarity, and caution must be exercised when applying these classifications in any study. Therefore, explicitation based on the aforementioned definitions and typologies is not easy to operationalize. In the following section, the major works which have been operationalized based on the definitions and classifications aforementioned are presented along with criticisms.

### ***Major Studies and Criticism***

As previously mentioned, so far there have been a lot of studies of explicitation. However, it should also be stressed that although a large number of studies have been conducted in explicitation, their findings are not conclusive and sometimes are contradictory.

In early studies on explicitation, almost all studies attempted to rely on the aforementioned definitions and classifications. Blum-Kulka's hypothesis has generally been confirmed by studies on translation between different language pairs, notably the investigation conducted by Linn Øverås (1998) and by Olohan and Baker (2000).

The first systematic corpus-based study on explicitation was conducted by Linn Øverås (1998) who used Blum-Kulka's Explicitation Hypothesis as the starting point. Øverås (1998:4) defined explicitation as the kind of translation process where implicit, contextually recoverable source text material is rendered explicit in the target text. To investigate explicitation shifts in translated texts, she employs a literary bi-directional English-Norwegian Parallel Corpus. She hypothesizes that the English and Norwegian translated texts are more cohesive than the source texts. Based on her findings, compared to the source text, there is a higher frequency of textual cohesion in both English translation and Norwegian translation. In addition, Øverås (ibid.) finds that the frequency of explicitation in the English-Norwegian translations is higher than that in Norwegian-English translation. On the one hand, Øverås (1998) argues that explicitation is confirmed as an inherent feature of the translation process by her study;

on the other hand, she remarks that explicitation might not always be universal since it can also be affected by other phenomena such as culture-bound translation norms and the stylistic preferences of source and target languages; with this in mind, explicitation should be studied further.

Øverås's (1998) work does not go without criticism. Becher (2010b: 12-13) points out that apart from a lack of contrastive study between languages, Øverås's study does not adhere to the definition of explicitation she proposes; Øverås counts and includes explicitation shifts found in her study which are definitely not co-textually recoverable. Becher (ibid.) argues that while recognizing the distinction between optional and translation-inherent explicitation, Øverås decides to include both types. Becher (2010a: 4) contends that, in principle, it would be possible that not a single case of explicitation found in Øverås's study belongs to the translation-inherent category and that her study does not therefore provide evidence for Blum-Kulka's Explicitation Hypothesis.

In 2000, Olohan and Baker also investigated the existence of explicitation. While Øverås (1998) uses two language-pair specific parallel corpora to investigate explicitation shifts, Olohan and Baker (2000) propose the study of translation universals regardless of the source language. They investigate the use of the optional reporting 'that' at the level of syntax in two monolingual comparable corpora, namely, the Translational English Corpus (TEC) and the British National Corpus (BNC). The comparable corpus used consists of texts selected from the BNC, which are as close as possible to the TEC in terms of size, time of publication and genre with over ninety percent of the corpus comprising of literary texts at that time (Olohan 2004: 94). Overall, they find that there is a higher level of that-connective in the Translational English Corpus than in the British National Corpus. Apart from the investigation of explicitation, the 'subconscious' nature of explicitation is also suggested in this study. Olohan and Baker (2000) hypothesize that translators may unintentionally use language in certain ways that increase the explicitness of the text without setting out to do so. Based on their findings, they suggest that the cognitive complexity involved in translation could explain the over-representation of the optional 'that' in translated texts.

Similarly to Øverås's work (1998), there have been various comments on the study by Olohan and Baker (2000). Becher (2011: 29) argues in terms of methodology, claiming that the corpus used (TEC) at that time was very imbalanced since the proportion of

fiction texts was very high compared to other genres; thus it can be described as a single-genre corpus and not a representative sample of translated English. In addition, Becher (2010a: 5) posits that while Olohan and Baker view observations as evidence of inherent, subliminal processes of explicitation in translation, another interpretation would be that their findings are simply the result of source language interference since many source languages in the corpus used, such as French, require a subordinator after reporting verbs; thus, the over-use of reporting 'that' in translated texts might be attributed to source language interference rather than to translation-inherent explicitation. Another interesting view is from Olohan (2004) who remarks on the notion of the 'subconscious process of translation' previously mentioned in Olohan and Baker (2000). Olohan (2004) realizes that the terminology employed to distinguish between deliberate and non-deliberate explicitation is loose and concludes that it is difficult to find the appropriate method to find evidence of the translator's motivation or lack of motivation when they employ conscious or unconscious strategies; however, she suggests that there is still much room for further possible refinement of the notion of explicitation to the extent that translators may make a deliberate decision to increase explicitness in a target text.

These two studies are considered early works in the systematic quest for explicitation. It is common to find scepticism about 'pioneer' studies, and such considered criticisms stand to benefit future work. To the present day, there has indeed been extensive research on explicitation, generating experimental data in many language pairs with the majority of studies based on a comparison of texts, often in corpora like those used in the aforementioned pioneer studies.

The pioneer studies on explicitation seem to support the existence of translation-inherent explicitation and can be considered a cornerstone of explicitation studies despite their criticisms. However, not all subsequent studies on explicitation based on corpora have been interpreted in the same way. Some translation scholars question the notion of explicitation as a universal characteristic of translation. In fact, House (2008: 6) strongly argues that the quest for specific translation universals is futile since as translation is an operation in language, general linguistic universals also apply by default to translation. House (*ibid.*: 11-12) explains that apart from an attachment to language universals, translations depend on language-pair specificity, directionality, genre-specificity and diachronic development. In pursuing her hypothesis, House



(2008) investigates explicitation in a corpus of children's book translations from English into German and German into English and claims that procedures of explicitation that hold for translations from English into German are not traceable in the opposite translation direction. House (2008: 12) argues that explicitation is not a universal feature of translation but a property which depends on the directionality of translation.

Learning from previous studies, Becher (2010a) realizes that many studies on explicitation suffer from the problems that (1) term 'explicitation' is not consistently and properly operationalized, and (2) most studies purport to offer data in confirmation of the Explicitation Hypothesis, taking the existence of translation-inherent explicitation for granted. Becher (2010b:3) writes, "Explicitation is observed where a given target text is more explicit than the corresponding text". The aim of his study is to test Blum-Kulka's Explicitation Hypothesis by investigating the occurrences of the deictic adverb in English-German translations where there is no equivalent expression in the English source text (Becher 2010a: 8). All texts under investigation are popular science texts. Becher (ibid.: 19) employs the filtering method which he considers the simple way to exclude obligatory, optional and pragmatic explicitations; nevertheless, he finds that it seems to be difficult to distinguish translation-inherent explicitations from optional explicitations at least in the domain of textual cohesion. Two causes are provided for his result: (1) the German preference for explicitness; and (2) source language interference (Becher 2010a).

Becher (2011) extends his research by investigating explicitation shifts on different levels (such as cohesive and denotative shifts) in an English-German business corpus. Based on his findings, he proposes a new explicitation hypothesis which claims that the occurrence of explicating shifts is uniquely determined by lexicogrammatical (such as the lexical inventory of the target language, and its syntactic properties) and pragmatic variables (such as the translators' willingness to take risks, their perception of their role as a cultural mediator, the communicative norms of the source and target language register) (ibid.: 19). He explains that the values of these variables are determined by the translation situation at hand (such as language pair, translator style, social factors). Furthermore, instead of resorting to the farfetched assumption of a "universal strategy inherent in the process of language mediation" (Blum-Kulka 1986: 21) or "subconscious processes of explicitation" (Olohan and Baker 2000: 141), he claims that

future research should seek to elaborate the above-mentioned lexicogrammatical and pragmatic factors, finding out how they differ across language pairs, translators and social contexts (Becher 2011: 219).

Based on all the aforementioned studies, it should be noted that although individual studies on explicitation could not yield conclusive answers, translation scholars can gain insight from previous research; Becher, for example, attempts to fill gaps he found in previous studies. So far, the broad hypothesis of explicitation has been refined by many studies, with translation scholars considering various factors affecting the manifestation of explicitation. Therefore, in contrast to Baker's (1996) initial proposal, many studies have shown that explicitation could be somehow affected by external factors, namely genre, language pairs, translators and contexts or cultures.

### ***Studies of Explicitation in Asian Contexts***

In terms of language pairs, undeniably the majority of the literature on explicitation (whether it is universal or not) has been drawn from an Anglocentric view where most supporting evidence comes from English and closely related European languages (Biel 2014: 109). To investigate whether a feature is universal however, Mauranen (2007: 45) argues that evidence from linguistically distant non-European languages is vital. Indeed, there have been a number of attempts at investigating explicitation in Asian contexts.

Chen (2004) explores cohesive explicitation of conjunctions in English-Chinese translations of popular science and information technology. He claims that, based on statistics presented in his study, translated Chinese in the genres under investigation tends to exhibit a higher level of conjunctive explicitness than either the source text or the comparable non-translated Chinese texts. In his study, he states that he finds various types of addition (as a realization of explicitation) which are not motivated by systematic differences nor any cultural gap; he argues that explicitated conjunctions in translated Chinese can be analysed and categorized according to the four types of explicitation presented by Klaudy (1996). Nevertheless, throughout his study, neither clear examples nor solid explanations are provided to support his claim; only the frequency of cohesive shifts is used to support the explicitation hypothesis.

Xiao (2010) also claims that, if the reported features of translational language are to be generalized as 'translation universals', the language pairs involved must not be

restricted to English and closely related languages. In addition, he states that since the translation universals proposed so far are identified on the basis of translational English, there is a possibility that such linguistic features are not universal but rather specific to English and genetically related languages. As such, he undertakes a project to investigate the features of translated Chinese in relation to non-translated Chinese. The ZJU Corpus of Translation Chinese (ZCTC) and the Lancaster Corpus of Mandarin Chinese (LCMC) are employed in this study. The genres covered in the ZCTC and LCMC are both literary and non-literary texts. Connectives are investigated to explore the explicitation phenomenon in native Chinese compared to translational Chinese. Xiao (2010: 18) finds that translated Chinese uses connectives more frequently than native Chinese which provides evidence in favour of the explicitation hypothesis. Nevertheless, this work provides no explanation regarding the types of explicitation found. The main focus of the study is again frequency of explicitation counted in the corpus as it only illustrates how often explicitation occurs not why and when it occurs. In this sense, the conclusive answer regarding the nature of explicitation in Chinese has yet to be given.

Again, these two studies on explicitation in Chinese cannot yield a conclusive result. Instead of providing qualitative analysis on explicitation, their studies mainly focus on the frequency of explicitation. Nevertheless, this could be considered an initial attempt to investigate explicitation in a language other than English and European languages and their studies could be later developed to find more explanation of explicitation in the Chinese language.

Explicitation has also been explored in the English-Korean language pair. Cheong (2006: 343-367) investigates target text contraction and target text expansion as realizations of implicitation and explicitation respectively based on a bidirectional parallel corpus of English-Korean. Cheong (ibid.) devises four measurement units, namely (i) word count rate, (ii) sentence initial connectives frequency rate, (iii) parenthesis frequency rate and (iv) bracket frequency rate to obtain an in-depth understanding of implicitation and explicitation (2006: 350). Apart from observing target text expansion and contraction and influence of language direction based on the aforementioned units, Cheong introduces his newly coined term ‘readership relations’ which he considers as a novel variable. According to Cheong (2006: 351), ‘readership

relations' refer to relative presuppositions-in-situation of the source text readership and target text readership.

Cheong (ibid.) states that his study offers counter-evidence against the prevailing belief in the universality of explicitation. Based on his statistical results, the translation of the English- into- Korean corpus leads to text expansion while the Korean into English leads to text contraction. Cheong (2006: 363) concludes that explicitation is not necessarily predominant or even present in all translations and the direction of translating language combination leads to a different target text expansion. Cheong states that his corpus consists of 517,607 tokens (299,630 in English and 218,979 in Korean); however, he does not exactly specify the size his sub-corpus (the number of words of English into Korean compared to that of Korean into English). Having said that, the aim to produce a bidirectional corpus could pose additional difficulties since material is seldom translated between two languages in equal quantities (Olohan 2004: 25).

### ***Different Approaches to Explicitation***

Translation scholars have attempted to find evidence of explicitation in corpora and to interpret findings in the light of translators' cognitive processes and language mediation effects. They have also attempted to explain explicitation based on non-linguistic factors. Currently, explicitation occupies a position at a crossroads, and involves studying translations as a product and translation as a process, and studying cognitive and sociocultural aspects of translation and even translator style (Kamenicka 2008: 1).

### **Explicitation as a Feature of Language Mediation**

Kruger (2012: 355-388) studies recurrent features or universals of English translation from Afrikaans. Apart from investigating the existence of three potential translation universals, namely explicitation, normalization and simplification, she focuses on the mediation effect in translated languages. She employs a cognitive approach and hypothesizes that universal features are the result of a mediation effect that is shared among different kinds of mediated language, rather than particularities of the translation process. Kruger (ibid.) defines mediated language as language that has undergone some kind of revision or rewriting after the original production with a particular purpose in mind. In this sense, Kruger (ibid.) clarifies that mediated language includes translation, editing, interpreting, subtitling and rewriting. This study uses a comparable corpus

(from four registers) consisting of a sub-corpus of English translations from Afrikaans, and of edited English texts and of unedited texts. Explicitation is operationalized and observed on the frequency of following linguistic features: (1) optional complementizer *that*; (2) full forms versus contracted forms; and (3) linking adverbs. Using the first and second operationalization, the study finds that the translated corpus demonstrates a lesser degree of explicitness than the other two sub-corpora. Furthermore, based on the third operationalization, the edited sub-corpus makes the least use of linking adverbs while the frequency of such is similar in the translated and unedited sub-corpora. The study then disconfirms the above hypothesis since there is almost no evidence of explicitation that is shared by translated and edited languages. Kruger (ibid.: 381) suggests that explicitation as observed in this study is a translation-specific effect since the effect is not shared by the two corpora of mediated texts.

It should be noted that the study of Olohan and Baker (2000) which was conducted a decade before still has an influence on Kruger's work (2012) in terms of the operationalization of explicitation. Furthermore, although no conclusive answers are yielded in this study, Kruger's (ibid.) attempt to investigate explicitation in both mediated and unmediated texts has brought the study of explicitation in new directions.

#### Explicitation in Relation to Readers

Saldanha (2008: 29) accepts that both cognitive constraints and textual linguistic conventions (such as communicative preferences across languages) can be seen as only two of the many factors that influence the translator's choice; however, there are still plausible factors influencing translators' decisions to employ explicitation. Instead of attempting to describe explicitation, Saldanha (2008) shifts the focus to explore motivations underlying translator's choices of explicitation. Saldanha (2008: 20) suggests that to fully understand the phenomenon of explicitation in translation, we need to go beyond only descriptions and dig deeper into translators' motivation for using explicitation as a strategy.

Saldanha (ibid.) argues that the studies mostly report on quantitative studies of explicitation and tend to rely on previous definitions without actually engaging with those definitions and their implications for their findings. She hypothesizes that explicitation should be recognized as a strategy that is associated with translators' assumptions about their readership and about their role as literary and cultural

mediators. Saldanha (2008) investigates the use of explicitation by two translators working from and into the same languages with a similar genre and in similar conditions and found that there is a more marked tendency towards explicitation in translations by one translator than in translations by the other. In the first instance, Saldanha considers that this is a result of the two translators' different stylistic preferences since the different use of explicitation could not be explained by different constraints in terms of languages, genres or cultural backgrounds.

To confirm this, Saldanha (*ibid.*) looks at these two translators' writing about how they do their translations and interviews both of them. Based on her findings, she concludes that the two translators employ explicitation differently because they have different concepts of their roles as intercultural mediators in relation to their audience. Saldanha (2008: 31) elaborates that it is possible to explain explicitation as a feature of audience design in which translators know how far they will go to meet the audience on its own terms and their willingness to align themselves occasionally with the source culture to present translated language.

This is an interesting hypothesis in explicitation since it illustrates that each translator is driven by their own desire to bridge the gap between source texts and translations; however, we should bear in mind that Saldanha's conclusion is inferred from indirect evidence. To some degree, it is difficult to measure the desire of translators. In addition, terms like 'desire' or 'willingness' are subjective; to support this hypothesis, in-depth analysis on extensive studies is necessary.

#### Explicitation and Risk

Pym (2005) models explicitation within a risk-management framework to provide a rational explanation of why explicitation phenomena might occur. Pym (*ibid.*) defines 'risk' as the probability of an undesired outcome. There are many risks in translation and a certain kind of risk which translators have to cope with is the risk of not being understood. According to Pym (2005), explicitation is one of the strategies to manage risk. Pym (2015: 73-76) indicates explicitation as a behaviour that is particularly risk-averse since translators presumably reduce the risk of end-user misunderstanding; the main task of translators is to ensure understanding between source language author and target language reader and to avoid misunderstandings at all costs.

If translators employ explicitation to manage risk, it is worth looking at instances when translators tend to employ explicitation to avoid risk. Pym (ibid.) explains that not all translators explicitate because they are all risk averse. Rather, risk aversion is a rational consequence of the kinds of situations in which translators work in certain cultures and within certain norms. Thus, in this sense, explicitation within the risk-management framework cannot be seen as a translation-inherent feature since the use of explicitation is determined by translator style and governing socio-cultural contexts.

This notion of explicitation has been embraced by Becher (2010: 20) who suggests that explicitation is worth employing to avoid misunderstandings in the communicative situation since if translation is too explicit, it might be a waste of energy but if the translation is too implicit, it might lead to communication breakdowns. Explicitation as a risk-avoidance strategy seems to be one of the plausible approaches to describe explicitation. However, to our knowledge, up until the present there have not been empirical studies investigating explicitation in relation to risk since all explanations are inferred from indirect evidence.

### **Unique Items Hypothesis**

The second translation feature that has a special status in this study relates to unique items. Compared to explicitation, this feature is much less researched. Nevertheless, it plays a vital role in the quest of features exclusive to translation, as we can see that, in recent years, this feature has gained attention from translation scholars and several studies have been conducted to investigate the so-called ‘unique items hypothesis’. This section starts with a definition of ‘unique items’ and criticisms thereof. It then presents major studies in this area.

#### ***Definition and Criticism***

Another concept proposed as a candidate translation universal (or, alternatively, a translation feature) is the so-called ‘unique items hypothesis’, which posits the under-representation of target-language specific items in translations. This concept is considered to be incompatible with the concept of explicitation and normalization, where translations are likely to exaggerate and over-represent specific items of the target language.

The notion of a unique items hypothesis was first introduced by Tirkkonen-Condit (2002). Unique items are considered linguistic elements which are unique in one

language as they do not have straightforward linguistic equivalents in other languages. They can be lexical, phrasal, syntactic or textual linguistic.

Tirkkonen-Condit (2004: 177-178) explains unique items, and the attendant unique items hypothesis, as follows:

Every language has linguistic elements that are unique in the sense that they lack straightforward linguistic counterparts in other languages. These elements may be lexical, phrasal, syntactic or textual, and they need not be in any sense untranslatable; they are simply not similarly manifested (e.g. lexicalized) in other languages. Since they are not similarly manifested in the source language, it is to be expected that they do not readily suggest themselves as translation equivalents, as there is no obvious linguistic stimulus for them in the source text.

The unique items hypothesis is operationalized by Tirkkonen-Condit (2002: 209) as follows:

[...] translated texts would manifest lower frequencies of linguistic elements that lack linguistic counterparts in the source languages such that these could be used as translation equivalents.

On this basis, Tirkkonen-Condit (2004) investigates the frequency of verbs of sufficiency and clitic particles in non-translated Finnish and translated Finnish in a corpus of fiction, on the one hand, and a corpus of academic texts, on the other. Verbs of sufficiency are selected because they constitute a lexical domain with no straightforward lexicalized equivalent in many Indo-European languages (ibid.: 178) while clitic particles are multifunctional and can be translated in many different ways. These items are investigated separately in the two corpora in anticipation of possible differences between the two genres (ibid.: 179). Tirkkonen-Condit (2004: 179-181) claims that the overall result of her study strongly confirms the unique items hypothesis both for verbs of sufficiency and clitic particles, especially in the fiction corpus, and the study also reveals the difference in the distribution of these items between these two corpora.

The concept of unique items has met with a number of criticisms. The major one is provided by Chesterman (2007: 5) who contends that although the unique items



hypothesis may be useful and could be tested in different conditions, it is not well defined and there seem to be many problems in the way it has been formulated. Chesterman (2007) argues that ‘unique items’ is not the right term to use, because it is too strong for various reasons. Chesterman (2007) holds that the concept is not clearly defined and raises a number of issues, namely whether the uniqueness is absolute or relative, whether the uniqueness is defined with respect to all other languages or a given source language, how to identify the unique items, whether the uniqueness is perceived linguistically or perceptually and whether the unique items are specific items to only translation. Chesterman (2007) posits that the unique items may be just a semantic void or lexical gap in the source language (or source-language lacunas) and concludes that Tirkkonen-Condit’s unique items should be postulated as a possible explanation for the under-representation of items, rather than as unique items (Chesterman 2007: 12). However, Chesterman (ibid.: 111-112) believes that in terms of research methodology, it is still possible to study ‘unique items’ if we adjust the methodology to fit what we are looking for.

### ***Major Studies***

To date, there have been many corpus-based studies attempting to verify the unique items hypothesis. Cappelle (2012) investigates verbs expressing manner of motion between original English (using data from the British National Corpus), English translated from French, and English translated from German (the latter two using data from the Translational English Corpus). The register is English prose. Bearing Chesterman’s (2007) criticisms in mind, Cappelle (2012) reformulates the definition of unique items but still asserts that his reformulation is consistent with Tirkkonen-Condit’s (2002) unique items. Cappelle’s (2012: 4) modified definition of the unique items hypothesis is that:

Translated texts manifest lower frequencies of linguistic elements that occur in one or more grammatical environments which lack linguistic counterparts in the source languages such that their out-of-context translation equivalents could also be used in these environments.

Cappelle (2012) conducts a contrastive-linguistic study of differences in verbs expressing manner of motion in English and French and he finds that manner-of-motion verbs are not found in French in the same range of grammatical environments as in

English; manner of motion verbs in his sense are considered to be unique items. Cappelle (ibid.) goes on to find that these manner of motion verbs are underrepresented in English translations from French. As such, he posits that this study confirms the manifestation of the unique items hypothesis (and at the same time contradicts the explicitation hypothesis).

In addition, Cappelle's study lends support to the notion of source language interference because its results also show that there are fewer manner-of-motion verbs in translations from French into English than from German into English. Nevertheless, the underrepresentation of manner of motion verbs in English translated from French, cannot be considered as a translation-inherent property since it does not occur in the same way in English translated from German (Cappelle 2012).

Another study that provides evidence in support of the unique items hypothesis is Vilinsky (2012). Vilinsky (ibid.) investigates the frequency of verbal periphrases in original Spanish and translated Spanish from English based on a comparable corpus of contemporary literary texts. Following Tirkkonen-Condit's concept and Chesterman's methodological suggestions on the unique items hypothesis, Vilinsky (ibid.) consults monolingual grammars and conducts contrastive studies between English and Spanish to find a set of periphrases which are unique to Spanish. Based on her data, Vilinsky concludes that her hypothesis, which is based on Tirkkonen-Condit (2002), that the target-language typical elements are likely to be underrepresented in translations, is confirmed.

Nevertheless, not all studies support the unique items hypothesis. One recent corpus-based study on unique items is Hareide (2016). Hareide (2016) investigates the manifestation of the unique items hypothesis by investigating the frequency of gerunds in original Spanish and translated Spanish from Norwegian. Hareide (2016) carries out a grammatical categorisation of gerunds between Norwegian and Spanish. Hareide (ibid.) discovers that not even one construction in Norwegian can straightforwardly express in the same way as the Spanish gerunds do and a wide variety of Norwegian syntactic and semantic constructions are translated into Spanish gerunds. She (ibid.) starts out by assuming that these Spanish gerunds can be unique items according to Tirkkonen-Condit's definition (2002, 2004). Nevertheless, her study disconfirms the unique items

hypothesis because the gerunds in the translated Spanish are overrepresented compared to those in original Spanish.

The results of studies of unique items in translation are again not very conclusive. Moreover, to our knowledge, no studies have yet been conducted to investigate the unique item hypothesis in any Asian language, including Thai.

## **Source-Language Shining Through**

### ***Definition***

Among the criticisms of studies of translation features is the fact that source languages are often not taken into consideration (Cappelle and Looock 2017). In line with Toury's (1995) law of growing standardisation and law of interference and Baker's (1995) normalization, Teich (2003: 145), however, investigates two "simple, intuitive assumptions about the nature of translations", namely source-language shining through and target-language normalization. According to Teich (2003: 145), source-language shining through occurs when translations tend to be more oriented towards the source language than the target language, while normalization occurs when there is an adjustment to properties of the target language resulting in the overuse of its features in the translations, and these two translation tendencies therefore make translations different from original texts in the same language.

### ***Major Studies***

The main goal of Teich's (2003) study is to investigate two specific features of translation, namely source language shining through and normalization. Teich (2003) employs a multilingual corpus consisting of original English and German and translated English and German to analyse the specific linguistic properties of translations. According to Teich (ibid.), this corpus is registerially controlled as it is compiled from only popular-scientific texts. Register analysis and contrastive studies of the properties (such as passive, transitivity and theme) of English and German are conducted based on Systemic Functional Grammar. According to Teich (2003: 207), the investigation of six features (including passive, agency and grammatical metaphor) shows that the English and German translations are different from the original English and German. In addition, with the discovery of the mixture of target-language normalization and source-language shining through, Teich (2003: 219) states that these two tendencies can co-occur. Moreover, Teich (ibid.) states that the different features are differently affected

by these two tendencies and to varying degrees, and some features are not affected by these two tendencies at all but by other factors. According to Teich (*ibid.*), when the target language has more options in its grammatical system, it enables the source language to shine through; however, when the target language has fewer options in a particular grammatical system, it has to compensate, but such compensation is not limited to normalization but can involve explicitation or compensation that does not result in normalization. As such, Teich (2003) argues that normalization cannot be tested only by comparing translations with comparable texts in the same language. It should be noted that Teich's (2003) interpretation mainly lies in the functional-grammatical and functional-contextual tasks that languages fulfil, rather than in cognitive processes.

Cappelle and Looock (2017) investigate the phenomenon of source-text interference in translations by taking typological differences into account. Cappelle and Looock (2017) adopt a similar approach to Cappelle (2012) but their study is not limited to individual languages like French and German, but encompasses different language families, namely Romance and Germanic. This study compares the use of the phrasal verbs in English translations translated from Romance languages and Germanic languages. The data are extracted from the fiction sup-corpus of the British National Corpus and the Translational English Corpus. Cappelle and Looock (2017) discover that phrasal verbs in English translated from Romance languages are underrepresented when compared with original English, while there is no difference in the use of phrasal verbs between English translated from other Germanic languages and non-translated English. In addition, English translated from the Romance languages contains fewer phrasal verbs than English translated from Germanic languages (*ibid.*). Cappelle and Looock (2017) assert that such results differ due to the language family to which the source language belongs. Romance languages, except Italian, do not have phrasal verbs, but make use of other structures such as prefixed and simplex verbs in French. The results show that translations from Romance languages do not use phrasal verb structures as frequently as English while translations from Germanic languages frequently use similar phrasal verb structures to English, thus it is possible to consider source-language family interference as a significant effect in translation (*ibid.*). Also, Cappelle and Looock (2017) dismiss 'normalization' and 'levelling-out' as translation universals because there is no strong evidence of these tendencies occurring in any language pair; in other words, English

translations of phrasal verbs from French are not fully normalized and underrepresentation of phrasal verbs does occur in English translated from Romance languages, but not from Germanic languages. Nevertheless, they hold that the unique items hypothesis could not be dismissed from their study. In addition to the study mentioned above, they also conduct a separate small-scale study to investigate how prefixed verbs and simplex verbs in French are translated into English. They (*ibid.*) find that French prefixed verbs are more frequently translated into English phrasal verbs than French simplex verbs because French prefixed verbs are conceptually close to English phrasal verbs. Cappelle and Looock (*ibid.*) conclude that this result reflects typological differences between source and target language, and that the nature of the source language shines through in translations.

In an attempt to investigate if there is evidence of Teich's (2003) 'source language shining through' in the English-Chinese language pair, Dai and Xiao (2011) examine the passive constructions in a mixed-genre Chinese monolingual comparable corpus and an English-Chinese parallel corpus. Based on their contrastive study of passive between Chinese and English, the use of passive in English is roughly ten times higher than that in Chinese (*ibid.*). After comparing translated Chinese with original Chinese, they show that translated Chinese contains a much higher incidence of passive constructions than original Chinese does. Dai and Xiao (2011) go on to explore their parallel corpus to see to what extent passive in Chinese translations is transferred from their English source texts. They find that 85% of passive constructions used in translated Chinese are influenced by the English source text.

Dai and Xiao's (2011) study thus provides supporting evidence for the interference of the source language in translation into Chinese (which is distant from non-European languages). It should be noted however, that Chinese and Thai belong to different language families—Chinese belongs to Sino-Tibetan whereas Thai belongs to Kra-Dai language family—so we cannot generalize from translation into Chinese to translation into Thai.

### **Studies of Translation Features in Legal Texts**

According to Biel and Engberg (2013), studies of language used in legal settings have gained a high degree of interest in recent years. Nevertheless, legal texts are still far less researched in translation studies than literary texts for example; and very few studies

have addressed the operation of translation features in translated legal texts. The exceptions include the work of Dorrit Faber and Mette Hjort-Pederson, who investigate explicitation in student and professional translations into English of a Danish law report (Faber and Hjort-Pederson 2009a, 2009b, 2013; Hjort-Pederson and Faber 2010), and Krogsgaard Vesterager (2017), who compares translations into Danish of a Spanish judgement completed by expert legal translators, on the one hand, and professional translators who do not specialize in legal translation, on the other.

Faber and Hjort-Pederson (see Faber and Hjort-Pederson 2009a: 108-109 and Hjort-Pederson and Faber 2010: 238-239) suggest one reason why explicitation might be expected to be rare in legal translation, namely that the addition of any information in target texts would be a risky strategy that could change the legal meaning and effect of the text, and it is therefore likely to be avoided. This argument is obviously particularly applicable to normative translations that, by definition, have legal effect. Conversely, they suggest a reason why explicitation might be expected, nonetheless, to appear in legal translations, especially those produced by novices: the linguistic complexity of legal texts makes them difficult to understand, and may trigger a process of mental inferencing which is ultimately evidenced in translated text. This, they claim, is consistent with Pym's (2005: 8-9) speculation that "the harder the source text, the harder the translator works, and the more likely they are to make their renditions explicit".

What is most interesting from our point of view are the points in source texts at which Faber and Hjort-Pederson expect mental and concomitant linguistic explicitation to occur. They include "nominalizations, passives, culture-bound terms and elliptical phrases" (2010: 241). Krogsgaard Vesterager (2017) follows suit in her study, ultimately concluding that explicitations of passives and nominalizations are rare in her data.

Faber and Hjort-Pederson's work is also of interest to us as they make a useful distinction between addition and specification. They use addition to denote a type of explicitation involving "the inclusion in the TT of extra lexical elements that either add or repeat meaningful elements" (2013: 44). Addition thus involves a quantitative increase in textual material. Specification, on the other hand, is qualitative in nature.

That is, “it adds meaning(s) by using lexical elements that are semantically more informative than the ST lexical elements” (ibid.).

However, these two works are not corpus-based. The use of corpora to investigate legal language and translation has been strongly supported by Biel (2010) and Biel and Engberg (2013). One of the contemporary corpus-based studies of legal translation is Biel’s (2014) work. Biel (2014: 118-119) investigates ‘textual fit’, the linguistic distance between translations and non-translations, of EU legislation translated into Polish compared with non-translated Polish legislation with the objective of enriching the understanding of EU translation and its effects on national legislation. Biel (2014: 118) operationalises the term ‘textual fit’ as “a linguistic distance between translations and non-translations of a comparable genre”. Biel (2014: 118-119) investigates ‘textual fit’ of EU legislation translated into Polish compared with non-translated Polish legislation. Biel (2014: 119) explains that this study adopts corpus findings on overrepresentation and underrepresentation of patterns as a starting point of textual fit analysis rather than starting with any hypotheses and also adopts the frequency-based approach and a broad view of phrasemes to examine how translated EU legislation departs from the target language conventions of national legislation.

Biel (2014: 137) claims that in her study, quantitative data serve as a basis for qualitative conclusions. Key generic patterns are identified with corpus-driven methods and also supplemented by patterns in her literature review of legal language (Biel 2014: 137). The four corpora employed in Biel’s (2014) project are (i) the English JRC-Acquis (source text); (ii) the Polish JRC-Acquis (authoritative and authentic translation); (iii) a self-built Polish Law Corpus; and (iv) the National Corpus of Polish.

Wordsmith is employed to find prominent patterns in the English Acquis and Polish Acquis corpora and the Polish Law corpus in order to identify lexemes which occur with a markedly higher frequency in the Polish JRC-Acquis corpus compared to the reference corpora.

After the key patterns are identified, they are categorized into two levels: macrostructure and microstructure. At the macrostructural level, textual fit is analysed at the level of generic text-structuring and grammatical patterns through the realization of the following linguistic features: (i) mental models of legal reasoning (if-then conditionals and related patterns); (ii) deontic modality (obligation and permission); (iii)

depersonalization and authorlessness (passive voice and impersonal patterns); (iv) logical relations between units of discourse (parataxis and hypotaxis); (v) qualifications and cohesion (adverbials and participles); and (vi) deixis (pointing devices) (ibid.: 137-143). At the microstructural level, textual fit is realized through term-embedding, term-forming and lexical collocations (ibid.: 223-264).

The selected generic patterns are then tested in the Polish JRC-Acquis corpus against the non-translated Polish Law Corpus and against a representative sample of the National Corpus of Polish, and the quantitative data are analysed qualitatively to investigate if the overrepresentation and underrepresentation of such patterns departs from the conventions of Polish national legislation and of general Polish respectively (Biel 2104).

Biel (2014) shows that the majority of the generic patterns of Polish translated law (Polish JRC-Acquis) are either underrepresented or overrepresented when compared to non-translated Polish legislation and the translations also contain structures which are atypical in Polish law. In her study Biel (2014) interprets the corpus results not only at the grammatical level, but more importantly, in the textualization patterns and cognitive structures of legal reasoning that are conventionally realized through such lexico-grammatical patterns.

### **Why Study Translation Features?**

Translation features have played a central role in debates in translation studies for several decades. Despite vast research on translation features based on corpora, there has not been a conclusive answer as to whether translation features manifest in all types of translation. In terms of terminology, this could be due to the vagueness of the concept of translation features, which means different scholars often operationalize this abstract notion in different ways. In terms of methodology, a corpus study sometimes has to rely on a number of subjective, sometimes even arbitrary decisions; for example, scholars have to choose which phenomena to exclude from analysis. For example, in classifying explicitation shifts, a number of assumptions have to be made, many of which are of an intuitive rather than objective nature (Becher 2011: 78). Thus, it is difficult to compare research results and replication of studies on translation features (Chesterman 2004: 44).



Nevertheless, there is strong support for studies on explicitation. Becher (2011: 78) states that the aforementioned criticisms and problems on explicitation are normal in all branches of applied linguistics (including translation studies). It is common that assumptions have to be made for practical purposes which cannot be objectively justified but without which an investigation of the phenomenon in question would not be possible (ibid.). Becher (ibid.) adds that current linguistic theory might not suffice to explain the manifestation of explicitation since many linguistic elements (such as modality) are never approached from the perspective of explicitation. In this sense, research into explicitation does not only benefit translation studies but also linguistics as a whole.

Chesterman (2010: 45-46) provides two reasons why translation universals should be studied. The first reason is in tune with what Becher (2011) argues above. Chesterman (ibid.) mentions that research on translation features will encourage translation scholars to develop better empirical research designs to work with large corpora and to think more about generating and testing hypotheses; this will help strengthen translation studies as an empirical discipline. The second reason is of pedagogical relevance: if it can be shown that there are certain features in translation, either negative or positive, we can educate students to improve the quality of their translation (ibid.).

Regardless of scepticism and criticisms involved, one thing which can be concluded from the review of studies on is that we learn more from the studies previously conducted by translation scholars from different times. It might not be possible to investigate translation features e.g., explicitation in translations from all times, language pairs and genres, but translation scholars are now realising the factors (namely, cognitive, social and linguistic) that should be taken account when designing their research on translation features. This is a promising sign since conclusive answers are yet to be found, but at least we are coming closer to this elusive goal.

Last but not least, translation features have gained currency in machine translation circles where translation features are used to explain differences in the performance of statistical machine translation engines trained on different data sets (Lembersky et al. 2012, 2013).

## Conclusions

As we have seen, previous studies have suggested that, on some occasions: translators make explicit in target texts information that is left implicit in source texts, or in comparable non-translated texts in the target language; that target language items that do not have obvious counterparts in the source language (and hence are ‘unique’ to the target language) may be underrepresented in translation; and that source-language interference can have a significant effect on translation. These tendencies have been viewed controversially as ‘universal’ features of translation, or as ‘translation inherent’, but recent studies have undermined the universalist agenda by showing how contrastive linguistic factors and typological differences affect explicitation, the distribution of unique items and the interference of the source language, respectively. It is thus possible to use the analytical categories of explicitation, unique items and source language shining through—as we do here—without subscribing to the idea of translation universals or translation inherence. And despite the difficulty in operationalizing some of these concepts, many authors, including Chesterman (2010), have argued that they are still worth pursuing, and authors in areas adjacent to translation studies have also found investigations into translation features of value to their work.

Moreover, even within a single target language, different domains and different target text functions may be associated with different translation strategies: translations of legislative texts that are not intended to have any legal effect, but merely to inform readers of the contents of the source text—corresponding to Cao’s (2007: 10-12) ‘informative’ translation—may, for example, tolerate explicitation more than ‘normative’ (ibid.) target texts, which are intended to be equally authentic to the source text (see related arguments in Dullion 2000 and Garzone 2000). There is still much to discover about translation in a variety of genres and a variety of language pairs. Writers like Teich (2003) and Biel (2014) provide good models of how to proceed with such research, and Biel (ibid.) also provides us with the concept of ‘textual fit’, which is sufficiently broad so as to allow a variety of comparisons between translations and other texts in the target language, without prejudging the explanations for the distance between translated and non-translated texts. In Teich (2003) and Biel’s (2014) work, as well as in many of the other sources presented in this chapter, explanations can come from language typology, contrastive linguistics, or genre studies, for example. In the next chapter, we turn to their, and ultimately our own, methodologies.

## Chapter 3: Methodology

### Introduction

The study of legal translation can benefit from corpus linguistics both for theoretical and practical use, and the way corpus techniques are applied is directly related to the type of corpus in use (Biel 2014). Based on research previously reviewed, we note that some studies can employ ready-to-use corpora. This is especially the case in the most researched languages like English, where researchers often use the Translational English Corpus and the British National Corpus, for example. Nevertheless, unlike English, Thai is a less researched language within the framework of corpus linguistics. Prior to commencing the work reported on in this thesis, we researched existing corpus resources in Thai to find out whether there are any corpora that are already available and which we could employ for this study. This chapter therefore starts with an overview of the corpus resources we managed to track down for Thai. From our research we concluded that the present study had to be based on specially constructed corpora. We thus provide here an explanation of how we created our parallel corpus of English-Thai legislative texts and our monolingual corpus of legislation originally drafted in Thai. We give the details of the design and compilation of the two corpora, and explain how we used Sketch Engine<sup>2</sup> to manage and interrogate them. Finally, we present the methodological considerations related to how we selected the features and linguistic framework used for analysis.

### Corpus Resources in Thai

Thai is generally considered a ‘resource-poor’ language (see, for example, Abdelali et al. 2014), although efforts have been made to make monolingual Thai corpora available to the research community through a variety of channels. One of the earliest such initiatives appears to have been ORCHID, a part-of-speech tagged corpus of technical texts compiled to support Natural Language Processing (NLP) research in Thai (Sornlertlamvanich et al. 1997). The Thai National Corpus (TNC) (Aroonmanakun 2007; Aroonmanakun et al. 2009) was subsequently designed as a general reference corpus of standard Thai and intended to be comparable to the written part of the British

---

<sup>2</sup> <https://www.sketchengine.eu/user-guide/user-manual/corpora/setting-up-parallel-corpora/#tab-id-2> (accessed December 2018).

National Corpus. It is freely accessible online through Chulalongkorn University. As of late 2017 it contained some 32 million words, over 1 million of which appear to be from legal texts, described as ‘regulation’. However, it is difficult to work out the exact composition of this subcorpus, which prevents us from integrating it into our study for the moment at least. The TNC is word tokenized—which is not a trivial matter, given that Thai does not use white space to mark word boundaries—but it is not part-of-speech tagged. The corpus is accessed through a web interface that allows concordance searches and provides information on the collocations and distributions of a given keyword.

Other monolingual Thai resources of which we are aware are composed of either untranscribed telephone conversations or transcribed children’s speech, or written texts harvested from websites, as in the case of the freely accessible 50-million word HSE Thai Corpus and the 82.8-million word ThaiWaC corpus, accessible through Sketch Engine. Finally, Thai was one of the South East Asian languages covered by the now-discontinued SEAlang Library Project, which provided access to dictionaries and monolingual and bilingual corpus resources.

In general, however, parallel corpora containing Thai are extremely rare. At least some data in Thai are included in the AMARA corpus of aligned educational video subtitles (reported on in Abdelali et al. 2014), and in Christodouloupoulos and Steedman’s (2015) ‘massively parallel’ corpus of the Bible, but we are not aware of any easily accessible general language parallel corpus including Thai, or any parallel corpus of legal language involving Thai. Nor are we aware of any Thai corpora, monolingual or bilingual, being used for translation studies research; the focus so far in Thai corpus development has instead been on the creation of corpora for use in general and computational linguistics (see, for example, Kawtrakul et al. 2002) and especially in NLP applications such as speech recognition (see above) and machine translation (e.g. AMARA).

Given the lack of suitable existing resources, the thesis has to rely on a specially constructed parallel corpus of legal texts in English and Thai, and a specially constructed monolingual corpus of Thai legal texts, both of which are described below.

## **Corpora in Use**

Legal corpora are employed in this thesis because the texts and translations under investigation in this study are legal texts. The two corpora used are: (i) a parallel corpus (English legal texts aligned with their Thai translations); and (ii) a monolingual comparable corpus of Thai legal texts (originals). These two corpora had to be compiled from scratch because, to date, as previously indicated, no parallel English-Thai legal corpus and no monolingual Thai legal corpus exist. Initially, we aimed to employ the TNC as a reference corpus; however, due to the difficulty mentioned above, we exclude it from our thesis. In the following sections, the general considerations influencing the design and compilation of specialized corpora are briefly outlined. The details of the corpora in use in the current research are then presented.

### ***Corpus Design for Special Purpose Text***

Kenny (2009: 59) states that best practice in corpus creation required designers to make informed decisions on the types of language they wish to include. In addition, when designing a special purpose corpus, many issues, such as size, number of texts, medium, subject, text type, authorship, language and publication date, have to be taken into consideration (Bowker and Pearson 2002). These and other issues as they impinge upon the design of the parallel and monolingual corpora built for this research are addressed below.

Another interesting criterion which has been mentioned with regard to corpus building is the notion of representativeness. Although many commentators argue that corpora should somehow be a representative sample of the language under investigation, the notion of representativeness is difficult to apply to textual data (Kenny 2009: 59). According to Kenny (2001: 107; drawing on Sinclair 2004), instead of seeking representativeness of corpora, researchers should focus on making their studies transparent by reporting text selection criteria as explicitly as possible. Therefore, the text selection criteria used in this study are also elaborated upon in detail below. The next section provides the details of how the two corpora are designed and compiled, starting with the English-Thai parallel corpus, followed by the monolingual Thai corpus.

### ***Parallel Corpus Design***

The self-built parallel corpus used in this research consists of the full texts of 25 international treaties in English and their translations into Thai. Based on the corpus management tool Sketch Engine, there are 172,739 words (tokens) on the English side and 222,556 words (tokens) on the Thai side. The corpus thus contains approximately four hundred thousand words (tokens). It should be noted that the number of Thai words counted by Sketch Engine is slightly different from that counted by Microsoft Word. The discrepancy is likely to be due to the different programs used to tokenize Thai words. We know that SWATH is used in Sketch Engine, but it was not possible to find out which program is used for counting Thai words in Microsoft Word. What we do know from Microsoft Support is that the tokenizer function used for Thai is embedded in the Word application by their software engineers, but how they have created the function and which program is used is Microsoft's proprietary knowledge and cannot be disclosed (personal contact with Microsoft Support).

The source texts are all international treaties entered into by Thailand and foreign countries, and came into force between 1950 and 2016. Only the English source texts are considered authoritative texts and the Thai translations have no legal force. The Thai translations are, however, vital to the civil servants and others who use them for information purposes. All Thai translations were completed and published by Thai government agencies. The original texts and translations were obtained in electronic format from the website of government ministries, departments and agencies. A manual search of all relevant websites was conducted specifically to locate international treaties and their translations. In total 26 treaties in English with Thai translations were sourced. One of these treaties, namely the United Nations Convention on the Law of the Sea, was excluded from the corpus for two reasons. First, this treaty contains around 90,000 words, while all other treaties sourced on the web contain around 5,000- 20,000 words. Such a high number of words would have skewed the corpus. Second, it was also impossible to convert the Thai translation of this treaty into a usable, machine-readable format.

After all texts are obtained, they were aligned semi-automatically in a format suitable for use with the Sketch Engine platform. Its detailed features are elaborated upon below.

### Language Direction

A decision was made early in this study to focus on English-to-Thai translation for two reasons. First, the author of this thesis is a native speaker of Thai. Second, while there may be plenty of Thai legal texts with English translations, to our knowledge, nearly all English translations are produced by translators who are native speakers of Thai. The translations are thus considered as inverse translations. Biel (2010: 7) contends that texts that have been translated by non-native speakers may contain some non-idiomatic expressions; therefore, to learn how things are expressed naturally in a language for specific purposes, it may be best to exclude non-native texts from the corpus. This is relevant to our research objective, which aims to investigate translation features in translated texts. Therefore, although it is difficult to obtain legal texts originally written in English and their Thai translations, it is still preferable to focus on this as the language direction for the current study.

### Size

It is often argued that a corpus used for the study of language for specific purposes (LSP) can be smaller than that typically used in language for general purposes (LGP) studies, because it is more difficult and time-consuming to obtain samples of specialized texts as opposed to general texts, and because specialized language represents a more restricted subset of natural language than LGP does (Bowker and Pearson 2002: 48). Well-designed corpora containing ten thousand to several hundreds of thousands of words have proved to be useful in LSP studies (*ibid.*). Likewise, according to Biel (2010), one of the major limitations concerning the use of legal corpora is their restricted availability, in particular for less researched languages.

These limitations are particularly acute in the case of our English-Thai parallel corpus of legal texts. The number of available English legal texts with Thai translations is quite limited because, while private legal texts such as private contracts and litigation documents are translated from English into Thai, it is generally impossible to access such documents for research purposes due to their confidentiality. (Based on the author's experience, the majority of private legal texts are prepared by private companies, such as international law firms, for their internal use; thus, the translations of these texts are not disclosed to the public.) International treaties are, however, publicly accessible in both the original English and in Thai translation, although their

numbers are limited. Nevertheless, for the purposes of the current study, 25 source texts were obtained, all of which gained legal force between the 1950s and 2016.

#### Genre, Authorship and Copyright

With respect to genre, it should be noted that all texts in the parallel corpus which are international treaties fall into the broad category of ‘legislative texts’ as defined by Cao (2010: 79). Regarding authorship, the English and Thai texts differ in their authorship, the former having been drafted by legal experts, the latter by professional translators, and, of course, in the translation condition—the English texts are originals, but the Thai texts are translations.

The original texts and translations were obtained from the websites of government agencies where copyright permission for the source texts and translations is not required because the use of the texts falls within an exception pursuant to section 32 (1) of the Copyright Act of Thailand B.E 2537 (1994), stipulating that the use of copyrighted work for research or study which is not for profit shall not be deemed an infringement. A full list of texts included in our parallel corpus, along with an indication of their provenance, is given in Appendix A.

#### Text Extent

Full texts are preferred to text extracts because linguistic features are rarely consistently distributed throughout texts and an extract can misrepresent the overall text (Kenny 2001: 110). In addition, in the case of language for specific purpose studies, if an extract is chosen at random, a part of the text that could be very interesting for the study might be accidentally eliminated (Bowker and Pearson 2002: 49).

#### Authenticity

As previously mentioned, only the original English legal texts in our parallel corpus are authoritative texts and the Thai translations, having no force of law, are considered as non-authoritative texts. We use the term ‘authoritative’ here to mean having the force of law (Šarčević 1997: 20). The term ‘authentic’, on the other hand, is quite problematic since it is defined differently in the context of law and linguistics; thus, to avoid any confusion in this research, the term should be clarified.

From the legal point of view, authoritative texts and authentic texts have the same status. To be authoritative, legislations must be adopted by the lawmaking body of the



particular country (Šarčević 1997: 20). Authoritative translations of legal instruments refer to translations which have been ‘authenticated’ in the manner prescribed by law and thus have the status of authentic texts (Šarčević 2012: 190). In these senses, all legally binding instruments, including authoritative translations, are also referred to as ‘authentic’ texts (Šarčević 1997: 20).

However, from the linguistic point of view, ‘authentic’ text refers to language that has been used by speakers or writers for the purpose of communication by real people in actual circumstances (Bloor and Bloor 2004: 5). Specifically, in the context of corpus linguistics, according to Bowker and Pearson (2002: 9), ‘authentic’ is one of the characteristics of a corpus and if a text is authentic, it means that the text is an example of real ‘live’ language and exemplifies genuine communication between people (as opposed to language invented by linguists to support particular arguments). Likewise, according to Stubbs (1996: 4), all corpus linguists must insist on the primacy of authentic data, as attested in texts that have occurred “naturally, without the intervention of the linguist”. The term ‘authentic’ is thus used very differently in legal and linguistic contexts.

The English language versions of the treaties are therefore considered ‘authentic’ in the sense that they are recognized as authoritative and as having legal force (Hermans 2007: 8; Šarčević 1997: 20). The Thai translations, on the other hand, are merely ‘informative’ (Cao 2007: 10). They serve as an explanation of the authentic version, but do not in themselves have legal force. Their lack of legal authenticity makes them no less important to their users however; on the contrary, their very existence attests to an acute need at the organisations where they are prepared and used. For example, the translations of the Charter of the Association of Southeast Asian Nations (ASEAN) and the United Nations Conventions contained in the parallel corpus were completed mainly by the Department of Treaties and Legal Affairs of the Thai Ministry of Foreign Affairs, and are used by even high-ranking officials, who are generally not fluent in English. Given that the translated texts were collected from a number of different sources, they represent the output of a range of translators. It is difficult to find information on the exact conditions under which the translations were created. Through our personal contact with a diplomatic officer in the Ministry of Foreign Affairs, we learned that translations of treaties are carefully prepared by a team of translators and the translations are revised many times until the final versions are obtained. However, to

our knowledge, we realize that there are no published accounts of contemporary translation activities in Thai ministries, departments and agencies, and there are no standard style guides or protocols that apply to translation across all Thai ministries, government departments and agencies. Table 3.1 illustrates the composition of the English-Thai parallel corpus.

	English Source texts	Thai translations
Size	172,739 words	222,556 words
Number of texts	25	25
Medium	Written	Written
Genre	Legislation	Legislation
Sub-Genre	International treaties	International treaties
Publication date	1950s to present	1950s to present
Authenticity	Yes	No
Authorship	Legal experts	Professional translators

**Table 3.1: English-Thai parallel corpus design**

After all texts are selected based on the above criteria, the texts need to be prepared before being loaded into Sketch Engine program.

#### Parallel Corpus Text Preparation

Legal corpora, like most specialized corpora, require the time-consuming alignment of the source texts and target texts (Biel 2010: 4). The majority of texts are in PDF format and the remaining are in .docx format (Microsoft Word). In fact, Sketch Engine supports multiple formats including doc, .docx, .htm, .html, .pdf, .ps, .tmx, .txt, .vert, .xm.

Nevertheless, it is much easier to align English texts with Thai texts when we convert files into Text Only format. After being converted, for the English original texts, there were very few conversion errors to be corrected. However, for the Thai translated texts, many conversion errors needed to be edited which was very time-consuming. After that, original texts and translations were semi-automatically aligned, as we did not have at our disposal a reliable process for fully automatically aligning English and Thai texts. Once the alignment was complete, the texts were ready to be loaded into Sketch Engine.

Upon ‘compilation’ of the English side of the corpus, Sketch Engine allows the user to specify a part-of-speech tagger that will be used to assign word class tags (for different forms of ‘noun’, ‘verb’, etc.) to individual word tokens in the corpus. We selected the

default modified English Tree Tagger for this purpose. It should be noted, however, that there is no comparable part-of-speech tagger available for Thai in Sketch Engine, and our efforts to source a part-of-speech tagger for Thai that could be implemented in Sketch Engine were not successful. The Thai side of our parallel corpus thus remains ‘untagged’.

### ***Monolingual Legal Thai Comparable Corpus***

The monolingual legal Thai comparable corpus consists of 1,173,485 words (tokens) (counted by Sketch Engine). The texts are all full, written, legislative texts enacted by a legislative body, and came into force between 1960 and 2016. The texts were obtained from the website of the Office of the Council of State.<sup>3</sup> A full list of texts included in the monolingual corpus is given in Appendix B.

Since this monolingual corpus is compiled to observe whether certain linguistic items in the translations are under- or over-represented when compared with non-translated texts in the same language, most of the criteria used in selecting the texts for the parallel corpus above are also applied here. Nevertheless, due to the differing nature of the two corpora, there are some differences in how certain criteria are applied, and these are discussed briefly below.

#### **Size**

As is predictable from comments made above on corpus size, and drawing on Bowker and Pearson (2002: 95) again, our monolingual corpus is bigger than our parallel corpus, as it is much easier to find original, monolingual Thai legal texts than it is to find parallel legal texts for the language pair English-Thai; thus, the number of words in our monolingual legal Thai corpus is more than double that of our parallel corpus. Again since the monolingual corpus is a specialized corpus, one million words for the comparable corpus are deemed appropriate.

#### **Genre, Authorship, Copyright and Text Status**

To make the monolingual corpus comparable with the parallel corpus, the same genre was selected. All texts in the parallel corpus are international treaties and their translations. Given the definition of a treaty specified in Article 2(1) (a) of the Vienna Convention on the Law of Treaties (1969), treaties fall into the category of primary

---

<sup>3</sup> [http://www.krisdika.go.th/wps/portal/general\\_en](http://www.krisdika.go.th/wps/portal/general_en)

legislation since they are passed by two or more governments. To select comparable texts which enjoy similar status, it is worth looking at the hierarchy of Thai legislation.

As mentioned in our discussion of Thai legal history in Chapter 1, continental law has had an influence on modern Thai law in many respects, including the concept of the hierarchy of law in Thailand. Based on the French legal system, Thai legislation in the present day can roughly be divided into five ranks, namely (listed from highest to lowest ranking): (i) constitution, (ii) acts, codes and emergency decrees, (iii) royal decrees, (iv) ministerial regulations and (v) municipal law (Uwanno 2000: 4).

The Constitution of Thailand is the supreme law of the country. No law can be in conflict with or contradict the constitution. The acts, codes and emergency decrees all have the same status although they are enacted by different bodies. These first two categories are considered primary legislation while the last three are considered subordinate legislation. To make the corpus comparable to the parallel corpus, primary legislation that came into force in the same period as the parallel corpus was selected. We inspected the lists of Thai laws on the website of the Office of the Council of the State and selected texts that constitute primary legislation. Our monolingual corpus thus contains one constitution, three codes, 110 acts and nine emergency decrees.

Since they are legislative texts, all texts are written by legal experts. In addition, the use of the texts is not bound by copyright under the condition of ‘fair use’ according to section 32 (1) of the Copyright Act of Thailand B.E 2537 (1994). Furthermore, all texts in the monolingual legal Thai corpus are authoritative and authentic in both the legal sense and linguistic sense.

Table 3.2 provides an overview of our monolingual corpus.

Size	1,173,485 words
Number of texts	123 texts
Language	Thai
Medium	Written
Genre	Legislation
Sub-Genre	Constitution, codes, acts and emergency decrees
Publication date	1960 to 2016
Authenticity	Yes
Authorship	Legal experts

**Table 3.2: Thai monolingual corpus design**

## Comparability

Both our legal corpora thus consist entirely of legislation, one of the four categories of legal text type recognized by Cao (2007). The texts in the Thai monolingual corpus are deemed to be comparable to the Thai translations in the parallel corpus, by virtue of shared values for attributes such as ‘Language’, ‘Medium’, ‘Genre’, and ‘Publication date’. They contrast with the Thai texts in the parallel corpus in having different values for ‘Authenticity’ and ‘Authorship’, as the Thai texts in this case are authentic and are written by legal experts.

The specific types of legislative text (called a ‘Sub-Genre’ here, following Biel 2014) contained in the monolingual corpus are also somewhat different from those in the parallel corpus, as international treaties are not drafted as original texts in Thai. The monolingual corpus instead contains other types of legislative texts. Care has been taken to select only primary legislation for inclusion in the monolingual corpus, so as to maintain parity of status between the texts included in both parallel and monolingual corpora. The monolingual corpus thus includes the Constitution of Thailand, codes, acts and emergency decrees, texts that fall into the two highest ranks of Thai legislative texts, as identified by Uwanno (2000: 4). It should be noted, however, that the primary legislation included in the Thai monolingual corpus differs from the international treaties in the parallel corpus in one other important way: while the former texts are imposed by the legislature, the latter are agreed between contracting parties. They thus differ in their genesis and function.

We also acknowledge here that while our parallel corpus contains informative translations in Thai with no legal force, our monolingual corpus of legislation in Thai does have legal force, and so comparability is somewhat undermined along the dimension of legal text function (Dullion 2000).

Given the different sizes of the two corpora, the monolingual corpus is also substantially larger than the Thai side of the parallel corpus. This is partly due to the relative ease with which monolingual corpora can be created, compared to parallel corpora (see, for example, Bowker and Pearson 2002: 95), but the larger size of the monolingual corpus allows us to have greater confidence in making claims about what is ‘normal’ in legislative texts in Thai. Quantitative comparisons between the Thai translations and the

Thai monolingual corpus are enabled by normalizing frequencies of occurrence of selected linguistic features, and thus reporting frequencies per million words.

As with other legal corpora, and especially parallel legal corpora, “which require time-consuming alignment” (Biel 2014: 94), our corpora are relatively small and are dominated by legislation. They are, nevertheless, substantial resources by Thai standards, and indeed by the standards of many other less researched languages (see Biel 2014: 93).

#### Text Preparation

Legal texts obtained from the Library section of the website of the Office of the Council of State are in html format. We first converted the texts from html to Text Only because Text Only format contains fewer errors than html. The texts were then loaded into Sketch Engine.

#### *Software Used: Sketch Engine vs ParaConc*

Unlike English, the Thai language is not a widely researched language in the field of corpus linguistics. Finding the appropriate software tool to build a Thai corpus was, therefore, a priority. To date, there are only few corpus management and analysis programs that can accommodate the Thai language, two of which are Sketch Engine and ParaConc.<sup>4</sup> These two software programs were tested for their usability in compiling and interrogating a Thai-English parallel corpus and a monolingual Thai corpus, as both programs are accessible under licence to staff and students at our university.

Sketch Engine is a corpus manager first developed by Adam Kilgarriff. It is a corpus query and management tool which is used not only by translators but also linguists, lexicographers, or anyone who is interested in studying how language behaves based on corpora. Sketch Engine is available to staff and students of DCU via a single-sign-on licence, which means it can be accessed online at [www.sketchengine.eu](http://www.sketchengine.eu) by remote users regardless of their location, as long as they have appropriate authentication.

ParaConc, a parallel concordancer developed by Michael Barlow, is a tool which can be used to produce instances of occurrences of a word or structure in the source texts and its equivalents in translation, or vice versa (Olohan 2004: 55). The version that we

---

<sup>4</sup> Available at: [www.athel.com](http://www.athel.com).

tested in a pilot study for the current research is ParaConc version 2003 (Barlow 2003), but it should be noted that the most recent version of ParaConc dates from 2009. To access ParaConc, program installation is required.

Given that the English language works well with these two programs, we built pilot corpora using these programs to compare how the programs work with the Thai language, in order to select the most appropriate tool for our main study. In the following paragraphs, details of how the two programs operate with the Thai language are briefly given to illustrate the similarities and differences between these two programs and to justify why Sketch Engine was chosen over ParaConc.

On the one hand, ParaConc and Sketch Engine are somewhat similar in terms of text preparation and corpus processing features. The Thai text preparation process for use with ParaConc and Sketch Engine is the same. To build a monolingual Thai corpus, for ParaConc, Thai text files must be saved in Text Only format and using the UTF8 encoding system before they are loaded into the program, but for Sketch Engine, more formats are acceptable as mentioned above. For the parallel corpus, both programs can align texts automatically, however, the degree of success also depends on the language pair. Chen (2004: 302) argues that problems in dealing with Asian languages generally in corpus-based translation studies are mainly related to parallel alignment and the lack of spaces between Asian characters. These problems also apply, of course, to the Thai language in particular. This is because Thai words are not orthographic words, and Thai sentences do not end with a delimiter. This makes not only word but also sentence tokenization difficult. This causes errors in automatic alignment, so to ensure that the texts are aligned correctly and to reduce any complications that might compromise the findings, texts are aligned semi-automatically when dealing with the English-Thai language pair. Once a pre-aligned sub-corpus is loaded into Sketch Engine or ParaConc in 'Text Only' format and using the UTF8 encoding system, the researcher can start processing the bilingual corpus. In addition, both programs offer all standard corpus-processing features, such as word and phrase search queries and parallel concordance searches.

On the other hand, Sketch Engine is considered more advantageous than ParaConc in many ways. First, it is easier to access Sketch Engine. Sketch Engine is accessible anywhere as long as the user's machine is connected to the Internet, while ParaConc is

accessible only on machines on which the program has been installed. Second, Sketch Engine is more robust than ParaConc. When testing ParaConc, we experienced some technical problems, for example, sometimes the program would not open or the corpus could not be saved properly. Such problems did not arise with Sketch Engine. Third, Sketch Engine can be used for the management of both monolingual and aligned, parallel corpora. The most important reason why Sketch Engine was selected for use in the current research, however, is that Sketch Engine comes with a useful tool for Thai, SWATH. SWATH is a Thai word tokenizer developed by Paisarn Charoenpornasawat (2003) and which was integrated into Sketch Engine in late 2016, specifically to facilitate the current study.<sup>5</sup> This tokenizer is considered necessary for this research because it makes Thai-word-based searches possible. If this tokenizer had not been incorporated in the program, Thai texts would have had to be tokenized before being loaded into the program. While it is possible to do this using a downloaded version of SWATH, this task is very time-consuming and requires use of the Linux operating system (as the Windows version did not appear to work properly in the trials we conducted). In addition, Sketch Engine also offers ready-to-use corpora in the Thai language, which can also be used as a reference corpus.

### ***Research Questions***

As previously mentioned, this thesis investigates certain linguistic features and their relationship to so-called ‘features of translation’, including the general distribution of items in translations compared to other texts in the same language, explicitation, source language shining through and the underrepresentation of unique items. Three overarching research questions have been formulated as follows:

1. How are instances of selected linguistic features translated from English into Thai in legislative texts?
2. Does the distribution of these features differ between translated Thai legislation and legislation originally written in Thai?
3. Can any differences in distributions of these features be explained by reference to explicitation, the unique items hypothesis, or source language shining through?

---

<sup>5</sup> The SWATH (Smart Word Analysis for Thai) word tokenizer is available for download from <http://www.cs.cmu.edu/~paisarn/software.html>



Each of these questions is reformulated in Chapters 4, 5 and 6 to instantiate the rather vague notion of ‘linguistic feature’ here with more precise values, namely ‘passive voice’, ‘deontic modality’ and ‘cohesive conjunction’.

To verify the existence or prevalence of translation features in legal Thai translations, first we need to operationalize the concepts of explicitation, the unique items and source language shining through used in this thesis. We also need to specify appropriate methods to select the linguistic features for analysis.

### ***Operationalization of Key Concepts***

Based on our literature review, we mostly follow the definitions of explicitation, unique items and source-language shining through as posited by Becher (2010), Tirkkonen-Condit (2002), and Teich (2003) respectively. The operationalization of these three features is elaborated upon in the following sections.

#### **Explicitation**

Explicitation has been a controversial idea in translation studies, with Saldanha (2008), for example, taking issue with some of the assumptions that scholars adopting the notion appear to make about the ‘implicitness’ of information in source texts, and the supposed increased ‘informativeness’ of explicitated target texts. Becher (2010a), likewise, criticizes research in the area for its lack of rigour, and is particularly sceptical of the notion that explicitation is a ‘translation-inherent’ phenomenon. He ultimately concludes that the explicitation hypothesis itself should be abandoned, “since it is unmotivated, unparsimonious and vaguely formulated” (2010b: 1). That other scholars have not abandoned the notion is clear from the number of studies that continue to be published on explicitation, among them a small number that deal with legal translation, and to which we return below. For now, it suffices to say that our approach to the subject follows Becher’s initial definition of explicitness as:

[...] the verbalization of information that the addressee would (most probably) be able to infer from the context, her world knowledge or from other inferential sources if it were not verbalized (Becher 2010a: 2).

We thus view ‘explicitation’ as the process of introducing verbalizations of such information to target texts. As already indicated, explicitation refers to a putative feature of translation whereby target texts are said either to represent information more

explicitly than their source texts, or to represent information more explicitly than similar, non-translated texts, in the target language. In the former case, studying explicitation requires access to a bilingual parallel corpus; in the latter case, it requires use of a monolingual, target-language comparable corpus. Studies exemplifying the former approach include Saldanha (2008) and Becher (2010a); those exemplifying the latter approach include Olohan and Baker (2000) and Cappelle (2012). In this thesis therefore we take a source-oriented approach in the first instance, comparing levels of explicitness in our source and target texts, followed by a comparison of the linguistic items in translated and non-translated Thai.

### Unique Items

As previously indicated, Tirkkonen-Condit (2004: 177-178) hypothesizes that ‘unique items’ are likely to be under-represented in translated texts compared to non-translated texts in the same language. As also previously mentioned, this ‘unique items hypothesis’ is criticized on a number of grounds (see, especially, Chesterman 2007). One relates to the very definition of ‘unique’ items. While the term suggests absolute exclusivity to the target language in question, in practice it is applied (by Tirkkonen-Condit and others) in cases where a target-language item does not have a ‘straightforward linguistic counterpart’ merely in one particular source language. By way of illustration, there might be no straightforward linguistic counterpart for some Thai linguistic items—for example certain passive markers in English—so if we are analyzing translations from English into Thai, we might consider such items as unique, and we might in turn expect them to be underrepresented in such translations (compared to native Thai texts). If, however, we are analyzing translations from Vietnamese into Thai, we are obliged to recognize the existence of a more-or-less straightforward linguistic counterpart of such items in the source language (see Keenan and Dryer 2007: 338; Prasithrathsint 2006: 116). In this case, the said items lose its status as a ‘unique item’.

Despite this and other issues with the definition and identification of unique items (Chesterman *ibid.*), we will continue to use the term here in its relative sense to mean ‘linguistic elements that occur in a given target language (Thai) but not in a given source language (English)’, for the sake of convenience—as corpus-based translation scholars are familiar with the term—and to maintain consistency with other studies that provide useful models for the current research (e.g. Cappelle 2012). Specifically, if we

were to find that the linguistic items are used in translation into Thai less frequently than in comparable original texts in Thai, we would have evidence to support the unique items hypothesis. If, however, we see more instances of such items in translated Thai than in non-translated Thai, this would contradict the unique items hypothesis.

#### Source language shining through

We follow the notion of source language shining through in translation proposed by Teich (2003). Teich (2003: 145) states that source-language shining through occurs when translations tend to be more oriented towards the source language than the target language. This is operationalized by comparing (i) frequencies of occurrence of selected linguistic features in the English source texts with corresponding features in the Thai translations (thus focusing on the translation relation) and (ii) frequencies of occurrence of selected linguistic features in the Thai translations and Thai original texts (Thai monolingual comparable texts). When discussing source language shining through, the first relation is in focus, while the second relation provides the grounds for comparison related to the specific properties of translations and Thai comparable texts.

#### *Selection of Features and Methods for Analysis*

The methodology employed in this research is inspired by Teich (2003), Becher (2010a) and Biel (2014). In these sources, a language-for-specific-purpose corpus is employed to investigate translation features in scientific texts, business texts and legal texts, respectively. The methodologies used in these three studies are different from each other. Therefore, it is worth providing a brief description of Teich's (2003), Becher's (2010a) and Biel's (2014) methodologies. Then we will justify the methods applied in the present thesis.

#### Teich's (2003) Methodology

Teich's (2003) study is hypothesis-driven. Teich (ibid.) first proposes two hypotheses: (i) translations are different from comparable texts because of source language shining through in translations, and (ii) translations are different from comparable texts because they try to be more rigid than the original texts in the same language. Then Teich (2003:3) states that a methodology for investigating the translation features in her study enjoys two properties, namely, it is theory-based and corpus-based.

In terms of theory-based properties, Teich (2003: 225) insists that her methodology is theoretically committed since its core component is a model of cross-linguistic variation

that is rooted in Systemic Functional Linguistics (Halliday 1985). Since the hypotheses concern not only the relationship between translations and comparable texts in the same language, but also that between the source texts and their translations, a contrastive analysis of source languages and target languages is required to ground the comparison (Teich 2003:143). Teich (2003) therefore conducts her contrastive analysis of English and German based on Halliday's (1985) Systemic Functional Linguistics to find the most prominent contrasts and commonalities between these two language systems. The contrastive typological observations thus derived are used as a basis for finding the most prominent features of English and German texts in the register of scientific writing, and six features are subsequently selected and analysed in the corpus.

In terms of corpus-based properties, the corpus compiled for Teich's project consists of four sub-corpora: English scientific texts, their German translations, German scientific texts and their English translations (Teich 2003: 147). The frequencies of occurrence of the six analytical features in the sub-corpora are compared based on three relations as follows: (i) the translation relation, the English source texts and German translations, and vice versa, (ii) the relation between original texts, the English original texts and German original texts, and (iii) the relation between the translations and original texts in one language (monolingual comparable texts) (ibid. 148). The first relation is in focus while the second and the third relations provide the grounds for comparison related to the specific properties of translations between English and German (ibid.). The results of the corpus analysis of the six features are then interpreted according to a model of cross-linguistic variation based on Systemic Functional Linguistics (ibid.). Teich (2003) asserts that the methodology of her study combines the strengths of corpus-based linguistic analysis with those of a theoretically well-founded functional-linguistic framework.

Becher's (2011) methodology

Becher's (2011: 122) study does not test a concrete hypothesis. Instead, it departs from the basic assumption that:

[...] every instance of explicitation and implicitation has a distinct cause, and the assumption of a 'translation-inherent' cognitive process of explicitation is not necessary [...]

to explain the frequent occurrence of explicitation in translation. His business corpus contains four parts composed of English texts, their German translations, German texts and English translations, and each part contains approximately 20,000 words. Given the relatively small size of corpus, Becher (2011: 78) reads through all corpus texts and manually identifies all explicating and implicating shifts based on assumptions of “an intuitive rather than objective nature”.

Becher (2011: 78) asserts that although his identification and classification of the explicating and implicating shifts are subjectively made, this is perfectly normal in studies in applied linguistics and also in translation studies. Assumptions made for practical purposes in this area sometimes cannot be objectively justified and without subjective decisions, an investigation of these assumptions would not be possible (ibid.). Regarding the linguistic framework that served as a guideline in his coding process, Becher (2011) sets up his own framework for identifying and classifying the shifts and which is loosely based on Halliday’s functional theory of language. He (ibid.) states that initially he intended to use the Hallidayan framework of Systemic Functional Linguistics (Halliday and Matthiessen 2004), but he later finds this framework to contain a considerable degree of vagueness and considers it hard to apply to his empirical data in a reliable way. He sets up the three shifts, namely interactional, cohesive and denotational shifts (which roughly correspond to Hallidayan’s three metafunctions of language), but he states that his framework draws heavily on the theory of formal semantics where the highly simplified – but practically useful – assumption is that language is primarily used as a means of representing the world.

#### Biel’s (2014) Methodology

Biel (2014: 118) defines ‘textual fit’ as “a linguistic distance between translations and non-translations of a comparable genre”. She stresses that textual fit is “a fuzzy concept” but operationalizes it using the notion of “atypical frequency (an underrepresentation and overrepresentation of lexicogrammatical patterns) and an atypical form of such patterns” (ibid.) Biel (2014: 118-119) investigates the textual fit of EU legislation translated into Polish compared with non-translated Polish legislation.

Her frequency-based approach means that Biel’s (2014) work is “purely data-driven” (ibid.: 119) rather than hypothesis driven and quantitative data serve as a basis for qualitative conclusions. Key generic patterns are identified with corpus-driven methods

and also supplemented by patterns in her literature review of legal language (Biel 2014: 137). Four corpora are employed in Biel's (2014) project: (i) the English JRC-Acquis (source text); (ii) the Polish JRC-Acquis (authoritative and authentic translation); (iii) the self-built Polish Law Corpus; and (iv) the National Corpus of Polish.

Biel (2014) employs software (Wordsmith) to find prominent patterns in the English Acquis and Polish Acquis corpora and the Polish Law corpus. Keyword and Wordlist functions in Wordsmith are employed to identify lexemes which occur with a markedly higher frequency in the Polish JRC-Acquis corpus than in the reference corpora, namely the Polish Law Corpus and the National Corpus of Polish. After the key patterns are identified, they are categorized into two levels: macrostructure and microstructure.

The selected generic patterns in the Polish JRC-Acquis corpus are then compared against the non-translated Polish Law Corpus and a representative sample of the National Corpus of Polish, and the quantitative data are analysed qualitatively, to investigate if there is overrepresentation and underrepresentation of such patterns when compared with the conventions of Polish national legislation and of general Polish respectively (Biel 2014).

We can see the methodologies used in these three studies are different. Teich's (2003) study is hypothesis driven and interpretation is mainly based on the functional-grammatical and functional-contextual tasks that languages fulfil, rather than on cognitive processes. Becher's (2011) study departs from assumptions about explicitation and relies heavily on formal semantics. Biel's (2014) study, however, is data driven, and she interprets corpus results not only at the lexico-grammatical level, but more importantly, from the point of view of textualization patterns and cognitive structures of legal reasoning that are conventionally realized through such lexico-grammatical patterns.

### Methods Used in this Study

In this section, we address the grounds we use to select the three linguistic features for analysis, namely passive, deontic modality and cohesive conjunction, and the linguistic framework used for our analysis. Broadly speaking, this work takes a lead from Biel's (2014) study where corpus evidence is used as a starting point of analysis. The broad methodology adopted in this study is somewhat similar to that in Biel (2014): we address both the relationship between our English legislative texts and their Thai

translations, and the relationship between the translations and comparable texts in Thai. However, in terms of methodology, it is difficult to adopt wholesale the corpus methods used in Biel's study.

Biel (2014) selects prominent features from her corpus, based on frequency of occurrence as ascertained using Wordsmith, and integrates these with the patterns identified as typical of legal language in her literature review. It is difficult to follow the same procedure in our corpus, however, due to the difficulties posed by written Thai. Although the SWATH program can be used to tokenize Thai words, tokenization is still difficult for the Thai language due to its complex nature. As such, after Thai words are tokenized by SWATH, the results yielded are not one hundred percent correct. If one tries to extract prominent features directly based on wordlists, for example, one finds a number of errors that could seriously compromise results of the research.

We thus look to Teich's (2003) study for alternative methods. We first assumed that it might be better to start by selecting prominent features from the contrastive analysis of English and Thai based on an appropriate linguistic theory. As already indicated, Teich (2003) selects analytical features after conducting a contrastive study between English and German based on Halliday's (1985) System Functional Linguistics. However, we again encounter problems due to the lack of contrastive resources for English and Thai. Given very limited resources, it is almost impossible to conduct a proper contrastive study between English and Thai and select salient features from the study. The majority of work that exists is mostly based on broad multilingual typological studies, for example Keenan's (1990) classification of world passive. We found no detailed, systematic, contrastive studies for the English-Thai language pair.

As such, we adopt Becher's (2011) justification regarding intuitive selection of the features for analysis. We select three linguistic features namely passive, deontic modality and cohesive conjunction because we assume that these features are good examples of features that manifest typological differences between English and Thai (in the case of passive voice), or that are particularly indicative of the genre in which we are interested (deontic modality and (conditional) cohesive conjunction are prominent features of legislative texts). We also suspect that these features might provide material traces of explicitation, source-language shining through and the under-representation of unique items. That said, the decision to study these linguistic features is not purely

intuitive, as we also select them based on our literature review. We note, in particular, that the linguistic features in which we are interested have been investigated by other researchers, for example Teich (2003), Dai and Xiao (2011) and Biel (2014). Further details of previous treatments of passive, deontic modality and cohesive conjunction, and of their realization in Thai, are provided in Chapters 4, 5 and 6, respectively.

Having selected relevant features characterising the lexico-grammar of English and Thai legislative texts, it is necessary to have an appropriate linguistic theory to guide the research. Like Becher (2011), our initial intention was to use Systemic Functional Linguistics to present linguistic phenomena and explain our results. However, as previously mentioned, there are only very limited existing resources that describe Thai based on Systemic Functional Linguistics. The majority of work is still based on traditional linguistics. We decided therefore to employ a neutral approach. Generally, we use standard linguistic treatments of Thai and English; but on occasions where there is a relevant Systemic Functional Linguistics treatment of a phenomenon in Thai, we refer to it as we deem appropriate.

Because two underlying relations will be examined: the relation between the translations and source texts and the relation between the translations and comparable texts in the target language, the selected features are investigated and compared in our different (sub-)corpora. The distribution of each selected feature in each sub-corpus is employed to investigate the relationship between source texts and target texts and to indicate if the features are overrepresented or underrepresented in translation when compared to non-translated comparable texts. The interpretation of results of the corpus analysis is carried out in two ways. As previously mentioned, traditional linguistics is used to explain the results. Nevertheless, the corpus findings are also interpreted according to patterns of the legal language addressed in our review of the literature on legal language.

Finally, we indicate here how we determine sample sizes in this research. Since we are working with a corpus, it is common to get hundreds or thousands of corpus lines for a search query and it is not usually possible to examine all corpus lines returned for a search. In such cases, it is acceptable to make observations based on a sample of concordance lines (Thomas 2016: 52). In Sketch Engine, the ‘Sample’ function can be



used to generate a random sample of concordance lines for a given search. For the purposes of this research, we determine the sample size as follows:

- If the number of results returned for a search is greater than 1,000 corpus lines, a one-tenth random sample is taken;
- If the number of corpus lines is between 301 and 1000, 100 random corpus lines are used as a sample; and
- If the number of the corpus lines is less than or equal 300, all lines are investigated.

## **Conclusions**

In this chapter we have described and justified our research design, providing details of the corpora used in our study and the methods we employ in selecting data for analysis. The detailed implementation of these methods and discussion of our findings and analysis is outlined in the following chapters on passive, deontic modality and cohesive conjunction.

## **Chapter 4: Passive Voice**

### **Introduction**

In this chapter we investigate the translation of passive voice in our parallel corpus, and whether there is evidence of explicitation, underrepresentation of unique items, or the source language shining through. The chapter starts with a description of the passive voice in Thai, which can be assumed to be unfamiliar to most readers. We provide details of corpus processing including the procedure followed in extracting data from the corpus. Then we present data from our corpus in the results and analysis section, and discuss their interpretation from the point of view of explicitation, unique items and source language shining through. The chapter ends with some observations on the limitations of the study.

### **The Passive Voice in Thai**

This section presents a short introduction to passive voice in Thai, based on general linguistic and typological studies (e.g. Keenan and Dryer 2007; Prasithrathsint 2003), and integrating the very limited literature on the Thai passive conducted within specific linguistic frameworks such as Systemic Functional Grammar (e.g. Patpong 2006).

The first thing to note about Thai passive is that some commentators express scepticism about its very existence. This is because passive is often considered a Western grammatical calque that came into Thai only in the early Bangkok period, less than 200 years ago (Prasithrathsint 2006: 116). Sudmuk (2003), for example, claims that passive structures in Thai are just ‘so-called’ ones influenced by English. Others, such as Prasithrathsint (2003, 2006), Kullavanijava (1974) and Iwasaki and Ingkapirom (2005), contend that a true passive voice does exist in Thai, with Prasithrathsint (2006) arguing that Thai passives are true passives since they are covered by Keenan’s (1990) classification of world passives. In addition, the alignment of certain passive forms with adversative meanings in particular has led some linguists to consider passive as a marginal phenomenon in Thai, and to deny that passive can occur with neutral or favourable meanings (Prasithrathsint 2006: 116).

Whether the passive in Thai is considered a ‘real’ or a ‘so-called’ one and whether it is used mostly to convey adversative meaning or not, Thai does have constructions used to

express passive meaning, on which Thai linguists agree to a large extent, and which are presented briefly in the following paragraphs.

Broadly speaking, there are two types of Thai passive structures: those with passive markers, also known as ‘overt’ passives; and those without passive markers, also known as ‘covert’ passives. In the following sections, the basic structure and common usage of Thai passive is first explained according to the general linguistic conventions.

### ***Passive with Passive Markers / ‘Overt’ Passive***

Overt passive constructions in Thai use one of three lexical items: ถูก /thuk/, โดน /don/ or ได้รับ /dai/+/rup/. These items are labelled differently by different Thai linguists: Patpong (2006: 273) describes them as ‘lexical verbs’ expressing the ‘receptive’ voice. Iwasaki and Ingkapirom (2005: 313) sometimes refer to them as ‘passive auxiliaries’ and sometimes as ‘passive markers’ (ibid.: 316; see also Prasithrathsint 2006). In the current study, the more neutral and broader term ‘passive marker’ is used.

The basic patterns in passive sentences using ถูก /thuk/ and โดน /don/ are identical. However, there is a slight difference in the passive pattern using ได้รับ /dai/+/rup/. Examples of use of the three passive markers are given below.

#### *ถูก /thuk/ and โดน /don/ passives*

In passives using the passive markers ถูก /thuk/ and โดน /don/, the patient-subject appears before the passive marker, followed by the agent (if present) and the lexical verb (Iwasaki and Ingkapirom 2005: 313-322). Auxiliary verbs (if present) come before the passive markers. The basic structure of this type of passive is thus:

*Patient-Subject + (Auxiliary Verb) + Passive Marker [/thuk/ or /don/]+  
(Agent) + Verb*

where parentheses indicate the optional nature of an element, and square brackets introduce passive markers from which one option must be chosen.

Examples 4.1(a-c) compare the active and passive versions of equivalent sentences, where the passive constructions use the ถูก /thuk/ passive marker:<sup>6</sup>

**Example (4.1)**

4.1a Active sentence:

Thai sentence:	ตำรวจจับขโมย
Gloss:	police arrest thief
English translation:	The police arrested the thief.

4.1b Passive sentence without agent:

Thai sentence:	ขโมยถูกจับ
Gloss:	thief /thuk/ arrest
English translation:	The thief was arrested.

4.1c Passive sentence with agent:

Thai sentence:	ขโมยถูกตำรวจจับ
Gloss:	thief /thuk/ police arrest
English translation:	The thief was arrested by the police.

Examples 4.1 (d-e) compare the active and passive versions with agent and the Thai auxiliary อาจ /aat/ (may).

4.1d Active sentence:

Thai sentence:	ตำรวจอาจจับขโมย
Gloss:	police may arrest thief
English translation:	The police may arrest the thief.

4.1e Passive sentence with agent and auxiliary

Thai sentence:	ขโมยอาจถูกตำรวจจับ
Gloss:	thief /aat//thuk/ police arrest
English translation:	The thief may be arrested by the police.

---

<sup>6</sup> Note that, unlike English, Thai is not inflected, and there is no morphological marking of tense, case or number. Linguistic glosses are thus kept to a minimum in these examples. The basic word order in Thai is Subject-Verb-Object.

The ถูก /thuk/ and the โดน /don/ passives have traditionally been used to express adversative meaning, in other words, to indicate that the action or event being described is unpleasant or undesirable for the patient-subject. ถูก /Thuk/, however, is increasingly also used in neutral situations (Prasithrathsint 1985; Iwasaki and Ingkapirom 2005).

Examples 4.2 (a-c) illustrate the usage of ถูก /thuk/ and โดน /don/ in adversative and neutral situations.

**Example (4.2)**

4.2a Adversative /thuk/:

Thai Sentence: แดงถูกแม่ตี

Gloss: dang /thuk/ mom hit

Translation: Dang was hit by Mom.

4.2b Adversative /don/:

Thai Sentence: แดงโดนแม่ตี

Gloss: dang /don/ mom hit

Translation: Dang was hit by Mom.

4.2c Neutral /thuk/:

Thai Sentence: เขาถูกแต่งตั้งให้เป็นหัวหน้าชั้น

Gloss: he /thuk/ appoint to be leader class

Translation: He was appointed as a class representative.

The literal meaning of ถูก /thuk/ as a transitive lexical verb is ‘to touch’ or ‘to hit on the point’ (Prasithrathsint 2006: 115). The meaning of โดน /don/ as a lexical verb is similar to that of ถูก /thuk/, but the โดน /don/ passive is more colloquial than the ถูก /thuk/ passive, and it indicates a strong sense of adversity (Iwasaki and Ingkapirom 2005: 317).

ได้รับ /dai+/rup/ passive

The ได้รับ /dai+/rup/ passive is frequently used in both positive and neutral situations. The word /dai+/rup/ is a transitive compound verb made up of ได้ /dai/ and รับ /rup/, both of which mean ‘receive’ and when it is used as transitive verb it can indicate either a favourable or an unfavourable situation (Iwasaki and Ingkapirom 2005: 317).

There are two passive patterns used with the passive marker ได้รับ /dai/+/rup/. The first is identical to that used with ถูก /thuk/ and โดน /don/, where the patient-subject appears before the passive marker, followed by the agent (if present) and the lexical verb.

The second pattern is normally used in formal language where ได้รับ /dai/+/rup/ can appear before what is normally a verb but which is nominalized using the prefixes การ /kan/ or ความ /khvam/ (Iwasaki and Ingkapirom 2005: 319).<sup>7</sup> The basic structure is:

*Patient-Subject + (Auxiliary Verb) + Passive Marker /dai rup/ +[/kan/ or /khvam/]+ Verb*<sup>8</sup>

As already indicated, the word ได้รับ /dai/+/rup/ is also used as a lexical verb meaning ‘receive’ or ‘obtain’. When this type of structure in Thai is translated into English, it can be interpreted in two ways, either as a passive construction or as an active construction with the lexical verb ‘receive’, as illustrated in Example 4.3.

**Example (4.3)**

Thai sentence: เขาได้รับการปกป้อง  
 Gloss: he /dai/+/rup/ /kan/ protect  
 Translation (passive): He was protected.  
 Translation (active): He received protection.

The ได้รับ /dai/+/rup/ passive is different from the ถูก /thuk/ and โดน /don/ passives in that it is not possible to convert an active sentence to its passive equivalent with ได้รับ /dai/+/rup/ while retaining the agent (Iwasaki and Ingkapirom 2005: 318-319).

**Passive without Passive Markers / ‘Covert’ Passive**

Thai passive constructions without passive markers, or ‘covert’ passives,<sup>9</sup> fall into two sub-types. In the first one, the patient-subject appears before the verb and is then optionally followed by a complement. The structure is as follows:

---

<sup>7</sup> The words การ /kan/ and ความ /khvam/ are classified as noun-forming prefixes: prefixes that attach to a non-nominal word to create nouns (Iwasaki and Ingkaphirom 2005: 28-29). The prefix /kan/ forms an activity noun with a verbal root (either with action verbs or mental-state verbs), and can also precede a verbal phrase while the prefix /khvam/ forms an abstract noun from an adjectival or verbal root and may appear before a verb phrase (ibid).

<sup>8</sup> Entire clauses can be nominalized in this way (i.e. verb + complement) (Iwasaki and Ingkaphirom 2005: 28-319).

*Subject-Patient+ Verb + (Complement)*

**Example (4.4)**

Passive sentence:

Thai:	บ้านนี้สร้างในปี 1994
Gloss:	house this build in 1994
Translation:	This house was built in 1994.

If the agent is important and needs to be referred to, the word *โดย* /doi/ meaning ‘by’ is added to indicate the agent.

*Subject-Patient+ Verb+ /doi/ Agent*

**Example (4.5)**

Passive sentence:

Thai:	หนังสือเล่มนี้เขียนโดยเจมส์ จอยส์
Gloss:	book this write /doi/ James Joyce
Translation:	This book was written by James Joyce.

According to Kullavanijava (1974: 196-197), this passive pattern is often heard in broadcasting and is found in written works that contain translations of English passive sentences. Iwasaki and Ingkapirom (2005: 316) agree that it has become common to use *โดย* /doi/ to translate English ‘by’ in contemporary written Thai and when *โดย* /doi/ appears, the passive marker can be left out.

It is thus possible that some Thai covert passives are the reduced form of overt passives, but to our knowledge, no studies to date confirm this possibility. In addition, at least one attempt has been made to categorize some Thai covert passives as middle constructions, with Inghongsa, Louie and Ruangjaroon (2016) suggesting that any Thai construction should be considered a middle construction rather than a covert passive when it has a subject interpreted as theme/patient along with an agentive verb and obligatory adverbial (ibid.: 220). The examples given in their paper are very limited however, and it would be premature to conclude that there is, indeed, a middle construction in Thai.

---

<sup>9</sup> According to Iwasaki and Ingkapirom (2005: 316), passive constructions that do not use passive markers are sometimes also called ‘verb passives’.

For the purposes of this study then, we recognize two types of Thai passive construction, based on the discussion above, namely:

- (i) the overt passive, which uses one of the three passive markers, ถูก /thuk/ ,โดน /don/ or ได้รับ /dai/+/rup/ and
- (ii) the covert passive, that is, the passive structure that does not use passive markers.

## Research Questions

In this study we are interested in both the relationship between source texts and translations ('equivalence' relations as defined by Biel (2014: 287)), and between target texts and other texts in the same language (Biel's 2014 'textual fit'). We set out to answer three main questions:

1. How are instances of passive voice translated from English into Thai in legislative texts?
2. Does the distribution of overt passives differ between translated Thai legislation and legislation originally written in Thai?
3. Can any differences in distributions be explained by reference to explicitation, source language shining through, or the unique items hypothesis?

## Research Procedure

In this section we outline how our corpora were queried. As indicated in previous chapters, Sketch Engine is used to manage and process our corpora.

Our empirical research begins with the identification in the English source texts in our parallel corpus of all instances of passive voice constructed using the lemma BE and a past participle. To do this, query (i) was created using Sketch Engine's Corpus Query Language (CQL):

- (i) [lemma="be"][tag="VVN"]

We wished to allow for cases where other constituents occurred between BE and the past participle, and so also used the following CQL searches:

- (ii) [lemma="be"][][tag="VVN"]



where [] indicates the presence of one token (e.g. ‘immediately’ in the string ‘X is immediately withdrawn’) between BE and the past participle, and

- (iii) [lemma=“be”][[]]{2}[tag=“VVN”]
- (iv) [lemma=“be”][[]]{3}[tag=“VVN”]
- (v) [lemma=“be”][[]]{4}[tag=“VVN”].

which allow for two, three or four tokens between BE and the past participle.

Queries (i) and (ii) yielded the vast majority of instances of passive voice in the English corpus, with 2,291 and 219 hits respectively. Hits in this case took the form of bilingual concordance lines.

Queries (iii), (iv) and (v) generated considerable noise, with only 20%, 9% and 1% of hits actually constituting passive patterns. We thus decided to focus entirely on hits for queries (i) and (ii). Using Sketch Engine’s random sample function, a sample of 230 instances (i.e. ~10%) was taken from the 2,291 hits for query (i). Likewise, a sample of 22 instances (again, ~10%) was taken from the 219 hits for query (ii).<sup>10</sup> Finally, all 252 concordance lines in our random sample were double checked to ensure that they were, indeed, passive constructions.

The next step in the procedure involved manually inspecting the Thai translation for each of the 252 instances of English passive voice in our random sample, to find out how they were translated. The findings of this analysis are presented in the section of Results and Analysis below.

As our research was concerned not just with how English passive is translated into Thai, but also with whether the distribution of the ‘unique items’ ถูก/thuk/, โดน /don/ and ได้รับ /dai+/rup/ differs between translated Thai and non-translated Thai, we also needed to search for instances of these lexical items on the target side of our parallel corpus and in our monolingual corpus of legal Thai. In this case, ‘simple’ searches were conducted for the items ถูก /thuk/, โดน /don/ and ได้รับ /dai+/rup/. Unsurprisingly, given its very

---

<sup>10</sup> Random sampling ensures that, in cases where there is too much data for the linguist to analyze manually, a more manageable, but unbiased sample can be analyzed instead. Given the lack of bias in the sampling strategy, there is no reason to believe that the 252 instances extracted are not representative of the approximately 2,500 instances of passive voice identified in the English source texts in the corpus.

colloquial nature, no hits were returned for the form โดน /don/ in either corpus, whereas 1,509 and 3,867 hits were returned for the forms ถูก /thuk/ and ได้รับ /dai+/rup/ respectively. As these forms are homographic with full lexical verbs or else may occur as part of certain compounds that were not accurately tokenized by SWATH, many hits for ถูก /thuk/ and ได้รับ /dai+/rup/ were false positives and needed to be weeded out. This was done manually, but the process was greatly assisted by the flexibility of the Sketch Engine user interface. The frequency of ถูก /thuk/ and ได้รับ /dai+/rup/ used as passive markers in the two Thai corpora was thus established. Given the difference in size between the two Thai corpora, the results for each were normalized by calculating frequency per million words. Further details are given in the section of Results and Analysis below.<sup>11</sup>

## Results and Analysis

In this section we first present the results from our analysis of the translation into Thai of instances of English passive voice. We then present the results from our comparison of the distribution of the passive markers ถูก /thuk/ and ได้รับ /dai+/rup/ in the Thai target texts in our parallel corpus, on the one hand, and our non-translated Thai legislative texts, on the other.

### *English-Thai Passives in Bilingual Parallel Corpus*

Table 4.1 shows how English passive clauses are translated into Thai in our parallel corpus. In roughly two thirds of cases (159 out of 252 instances or 63% of the time), Thai translators translate English passives using Thai passives. In such cases, the covert passive is used almost twice as often (105 out of 252 instances, or 42% of the time) as the overt passive (54 instances, or 21%). The second most common technique is to convert the English passive clause into a Thai active clause, which happens just over one quarter of the time (66 cases, or 26%). A relatively small number of other techniques are also employed, including omission of the passive clause or the shifting of the passive clause to a lower rank, such as a nominal group. There are 27 such instances in our data, accounting for 11% of the 252 cases studied. A more detailed, qualitative analysis of the data depicted in Table 4.1 follows.

---

<sup>11</sup> In this case, the intention was to conduct a purely quantitative analysis of the passive markers in question, once identified, and no further sampling to reduce the data set was necessary.

Translation techniques	Number of occurrences	Percentage
<b>1. English passive to Thai passive</b>	<b>159</b>	<b>63</b>
• Covert passive	105	42
• Overt passive	54	21
/thuk/+verb	23	
	<i>[9(negative),14(positive or neutral)]</i>	
/dai/+rup/+nominal group	23	
/dai/+rup/+ verbal group	8	
<b>2. English passive to Thai active</b>	<b>66</b>	<b>26</b>
• Full Thai active clause (agent known in the source)	26	
subject+verb+complement	21	
/hai/+subject +verb+complement	5	
• Active without subject (agent unknown)	36	
verb+compliment	26	
/hai/+verb+complement	10	
• Active with added subject (agent unknown)	4	1.6
<b>3. Others (omission, change clause to phrase)</b>	<b>27</b>	<b>11</b>
<b>Total</b>	<b>252</b>	<b>100</b>

**Table 4.1: Three main techniques employed by Thai translators when translating English passives**

#### **English Passive to Thai Covert Passive**

As previously mentioned, 105 instances (or 42%) of English passives were translated into Thai covert passives. The high frequency of this passive pattern in the Thai translations is consistent with Kullavanijava's (1974: 196-197) and Iwasaki and Ikaphirom's (2005) claim that Thai covert passives are characteristic of translations from English. In the absence of a comparative study of the distribution of covert passives in comparable translated and non-translated Thai texts, however, it is not possible to establish a causal relationship between the translation condition and the prevalence of covert passives in translated Thai.

The Thai covert passive in our parallel corpus was found in both patterns described above, namely

*Subject-Patient+ Verb + (Complement)*

and

*Subject-Patient+ Verb+ /doi/ Agent*

as highlighted in Examples 4.6 and 4.7 respectively.<sup>12</sup>

#### **Example (4.6)**

English source text:

*...a certified copy shall be sent*, through diplomatic channels, to each of the States Members of the Hague Conference on Private International Law...

Thai translation:

...สำเนาเอกสารดังกล่าวจะส่งตามวิถีทางการทูตไปยังแต่ละรัฐสมาชิกของที่ประชุมกรุงเฮกว่าด้วยกฎหมายระหว่างประเทศแผนกคดีบุคคล

Back translation (gloss):

*copy document such will send* through channel diplomat to each state member of the conference Hague on law international department case private...

#### **Example (4.7)**

English source text:

*Any organization* referred to in paragraph 1 above which becomes a Party to this Convention without any of its member States being a Party *shall be bound by* all the obligations under the Convention.

Thai translation:

องค์การใดที่อ้างถึงในวรรค 1 ข้างต้นซึ่งเข้าเป็นภาคีอนุสัญญานี้โดยที่ไม่มีรัฐสมาชิกใดของตนเป็นภาคีต้องผูกพันโดยพันธกรณีทั้งปวงภายใต้อนุสัญญานี้

Back translation (gloss):

*organization any* that refer in paragraph 1 that become party convention without state member as party *shall bind /doi/ obligation* all under convention this

Based on our findings, the verbs used in this pattern are very varied, but none of them is particularly associated with adversative situations.

---

<sup>12</sup> Note that only the instance of passive voice under discussion is highlighted in these examples, and the concordance lines may contain further instances of passive voice that are not discussed here.

## English Passive to Thai Overt Passive

As already indicated, 54 instances of the overt passive were found in our Thai translation data (accounting for 21% of our original 252 cases). Of these, 23 are passive constructions with the passive marker ถูก /thuk/, of which 9 instances express adversative meaning while the remaining 14 express either positive or neutral meaning. Examples 4.8 and 4.9 illustrate the ถูก /thuk/ passive used to express adversative and neutral meaning respectively.

### Example (4.8)

English source text:

*An alien* lawfully in the territory of a State Party to the present Covenant *may be expelled* therefrom only in pursuance of a decision reached in accordance with law.

Thai translation:

คนต่างด้าวผู้อยู่ในดินแดนของรัฐภาคีแห่งกติกานี้โดยชอบด้วยกฎหมายอาจถูกไล่ออกจากรัฐนั้นได้โดยคำวินิจฉัยอันได้มาตามกฎหมายเท่านั้น

Back translation (gloss):

*alien* who live in territory of state party this lawfully *may /thuk/ expel* from state that by decision from law only

### Example (4.9)

English source text:

Any such plan should be transmitted to the Secretariat as soon as *it has been developed*.

Thai translation:

แผนใด ๆ ดังกล่าวควรจะถูกส่งให้กับสำนักเลขาธิการโดยเร็วที่สุดหลังจากที่ถูกพัฒนาขึ้น

Back translation (gloss):

plan any such should will /thuk/ send to secretariat as soon as /thuk/ develop

Our data thus appear to confirm the claim, reported above, that ถูก /thuk/, although traditionally used in adversative situations, is now increasingly also used in neutral

contexts. Such observations may indicate the increasing grammaticalization of forms such as ถูก /thuk/.<sup>13</sup>

The ได้รับ /dai/+/rup/ passive appears in the other 31 cases where an English passive is translated by a Thai overt passive. In 30 out of 31 instances it is used to convey a positive or neutral meaning.

There are 8 instances where ได้รับ /dai/+/rup/ is followed by a verb, and 23 instances where it is followed by [/kan/ or /khwam/] + Verb. Example 4.10 illustrates the use of the ได้รับ /dai/+/rup/ passive in the former pattern while Example 4.11 illustrates the latter.

#### **Example (4.10)**

English source text:

One competent authority *shall be designated* to receive the notification in case of a State of transit.

Thai translation:

ศูนย์ประสานงานหนึ่งแห่ง *ต้องจะได้รับแต่งตั้ง* ให้รับการแจ้งในกรณีเป็นรัฐที่ถูกนำผ่านแดน

Back translation (gloss):

centre co-ordinate one *shall will /dai/+/rup/ designate* to receive notification in case be state which transits

#### **Example (4.11)**

English source text:

...and *their defence counsel*, as well as persons participating in the investigation, *are protected* against all ill-treatment or intimidation...

Thai translation:

และ *ทนายของบุคคลดังกล่าว* ตลอดจนบุคคลอื่นๆ ที่มีส่วนร่วมในการสอบสวน *ได้รับการคุ้มครอง* จากการปฏิบัติที่เลวร้าย หรือการข่มขู่

Back translation (gloss):

...and their defence counsel as well as person other who participate in investigation */dai/+/rup/ /kan/ protect* from treatment bad or intimidation

---

<sup>13</sup> Put simply, grammaticalization concerns the development of lexical items into markers of grammatical categories (Bisang 2017).

### English Passive to Thai Active

As indicated above, the second most common technique used by the translators when translating the English passive to Thai is to convert English passive clauses into Thai active clauses, which occurred in 66 (26%) out of 252 instances. The resulting Thai active clauses can be divided into three sub-categories as described below.

#### Translation to Thai Full Active Clause

In 21 cases where an explicit agent in the English clause was available, translators converted the English passive clause into an active Thai clause with SVO word order as illustrated in Example 4.12.

##### Example (4.12)

English source text:

...no further communication by any state Party shall be received under the present article after *the notification of withdrawal of the declaration has been received by the Secretary-General*, unless the state Party concerned has made a new declaration.

Thai translation:

ภายหลังเลขาธิการได้รับเรื่องแจ้งขอถอนประกาศแล้ว จึงไม่ต้องมีการรับคำร้องเรียนอื่นใดโดยรัฐภาคีภายใต้มาตรานี้อีก  
ยกเว้นแต่รัฐภาคีที่เกี่ยวข้องได้ทำประกาศขึ้นใหม่

Back translation (gloss):

...after *secretary general receive the notification of withdrawal* so must not have receipt of communication any by state party under article this unless state party that relate make declaration new

In a further five instances translators not only converted English passive to Thai active, they also put ให้ /hai/ at the beginning of the sentence as in Example 4.13

##### Example (4.13)

English original text:

The frequency of subsequent submission of information required under this Article *shall be determined* by the Conference of the Parties serving as the meeting of the Parties to this Protocol...

Thai translation:

ให้ ที่ประชุมสมัชชาประเทศภาคีอนุสัญญาฯ ที่เป็นการประชุมของประเทศภาคีพิธีสารนี้ กำหนด ระยะเวลาในการนำเสนอ  
ข้อมูลครั้งต่อไป ที่กำหนดภายใต้พิธีสารนี้

Back translation (gloss):

*hai* meeting conference party serve as meeting of party protocol *determine* period of time to submit information subsequent require under protocol this

A detailed discussion of ให้ /hai/ is beyond the scope of this thesis. Suffice it to say here that the placing of ให้ /hai/ at the beginning of sentences like that in Example 4.13 creates a causative construction (Iwasaki 1998) in which one entity, the ‘trigger’ (ibid.), causes, enables or obliges another entity, the ‘instigator’ (ibid.) to carry out some action. In Example 4.13 the trigger is only implied (but is understood to be the law itself), while the instigator is ‘the Conference of the Parties...’. Here ให้ /hai/ also expresses a degree of obligation similar to that of the modal ‘shall’ in the English source text, and the analysis of the translation of passive voice overlaps with the analysis of modality, which we treat in depth in Chapter 5.

### Translation to Thai Active Clause without Subject

In 26 cases where there was no explicit agent in the English passive clause, the translators translated the English passive to a Thai active clause without a subject using the pattern: verbal group + complement, as illustrated in Example 4.14.

#### Example (4.14)

English source text:

The persons concerned *shall be informed* of these rights before or at the latest at the time the decision is rendered.

Thai translation:

ต้องแจ้งสิทธิเหล่านั้นแก่บุคคลที่เกี่ยวข้องก่อน หรืออย่างช้าที่สุดในเวลาที่ออกคำตัดสินนั้น

Back translation (gloss):

*Shall inform* right those to person related or latest in time issue decision that

In a further 10 instances the translators used a ให้ /hai/ causative construction, but unlike in the instance referred to in Example 4.13 above, in this case the translations did not express an instigator, but followed the structure: /hai/ + verb + complement, as illustrated in Example 4.15.

#### Example (4.15)

English source text



Any dispute arising from the interpretation of the Charter *shall be settled* in accordance with the relevant provisions in Chapter VIII.

Thai translation

ให้ระงับข้อพิพาทใดๆ ที่เกิดจากการตีความกฎบัตรตามบทบัญญัติที่เกี่ยวข้องในหมวด 8

Back translation (gloss):

*/hai/ settle dispute* any arise from interpretation according to provision related in chapter 8

The omission of an instigator, and hence of a grammatical subject, in the Thai translation in Example 4.15 (and which makes the gloss back translation read like an imperative) may be linked to the absence of an agent in the passive construction used in the source text.

### Translation to Thai Active Clause with Added Subject

Third, there are four instances in our data where there is no agent specified in the English passive, but the translator added an agent as subject in the Thai translation, as illustrated in Example 4.16.

#### Example (4.16)

English source text:

*External parties may be invited to ASEAN meetings* or cooperative activities without being conferred any formal status, in accordance with the rules of procedure.

Thai translation:

อาเซียนอาจเชิญภาคีภายนอกให้เข้าร่วมการประชุมหรือกิจกรรมความร่วมมือโดยมิต้องกำหนดให้สถานภาพอย่างเป็นทางการใดๆ ตามกฎว่าด้วยขั้นตอนการดำเนินงาน

Back translation (gloss):

*asean may invite party external* to join meeting or activity without confer status formal any in accordance with rule of procedure

In these cases, we assume that the translator has verbalized ‘information that the addressee would (most probably) be able to infer from the context, her world knowledge or from other inferential sources if it were not verbalized’ in line with Becher’s definition of explicitation given above (from Becher 2010a: 2, and repeated here for convenience), and so these four cases constitute genuine instances of explicitation. We

return to this issue again below, after results from our monolingual comparable corpus investigations have been presented.

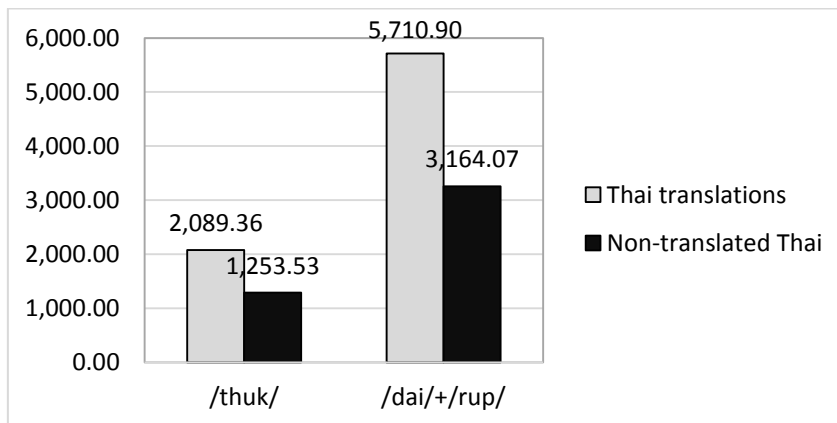
### ***Translated vs Non-Translated Thai***

Table 4.2 shows the distribution of **ถูก** /thuk/ and **ได้รับ** /dai+/rup/ used as passive markers in the Thai translations in our parallel corpus and in non-translated texts in our monolingual Thai corpus.

<b>/thuk/</b>			
Corpus	Number of occurrences	Corpus size (tokens)	Frequency (per million)
Thai translation	465	222,556	2,089.36
Non-translated Thai	1,471	1,173,485	1,253.53
<b>/dai+/rup/</b>			
Corpus	Number of occurrences	Corpus size (tokens)	Frequency (per million)
Thai translation	1,271	222,556	5,710.9
Non-translated Thai	3,713	1,173,485	3,164.07

**Table 4.2 Occurrences of passive markers /thuk/ and /dai+/rup/ in Thai translation and comparable non-translated Thai**

Figure 4.1 provides an alternative visualization of the relative frequencies of **ถูก** /thuk/ and **ได้รับ** /dai+/rup/ in the same two corpora.



**Figure 4.1: Frequency per million words of passive markers /thuk/ and /dai+/rup/ in Thai translated and Thai non-translated legal corpora**

Figure 4.1 suggests that ถูก /thuk/ and ได้รับ /dai+/rup/ are overrepresented in legislative translations into Thai compared to comparable legislative texts originally written in Thai. The implications of this finding for the unique items hypothesis are discussed below.

## **Discussion**

### ***Relation between Source and Target Texts***

Given the typological differences between English and Thai discussed above, the passive voice remains an intriguing issue, if the passive voice is considered a ‘marginal phenomenon’ or even, somewhat derogatively, a ‘foreign import’ in Thai (Prasithratsint 2006).

This study began with the observation that explicitation might be expected to be a feature of legal translation into Thai if translators (a) sought to avoid passive voice and subsequently replaced it by active voice, on occasion thereby making explicit agents in the target language that were not explicit in the source text or (b) if they preserved passives in translation into Thai, but in doing so added Thai passive markers that in turn made explicit the favourable, neutral or adversative nature of the situation described.

Even if passive voice is a relatively new phenomenon in Thai, it is also argued that Thai has by now well-established markers, namely ถูก /thuk/, โดน /don/ and ได้รับ /dai+/rup/, which indicate passive voice, but may also signal adversative, neutral or positive meanings (Prasithratsint 2003, 2006). These markers have no obvious formal counterparts in English, the source language under consideration here. While English marks passive voice using the auxiliary verb BE and the past participle of the lexical verb, it does not have specific grammatical or lexical words that indicate, for example, both passive voice and adversative meaning in the same token. Relative to English, ถูก /thuk/, โดน /don/ and ได้รับ /dai+/rup/ thus appear to be ‘unique items’ in the sense first put forward by Tirkkonen-Condit (2002), (and expanded upon the following part). What is also interesting about these items from our point of view is that if they are used in translation into Thai, in instances where the adversative, positive or neutral nature of the situation or action described in English is not explicitly verbalized but may be inferred (again as per Becher’s (2011) definition), then the presence of these items in Thai could

constitute evidence of explicitation, whereby, for example, the adversative nature of an action in English is made explicit in the Thai passive marker ถูก /thuk/ or โดน /don/.

Nevertheless, our finding is that our data do not support the expectation that passive is avoided in translation into Thai. In 63% of cases, English passives are translated by Thai passives. In slightly more than 26% of cases, however, English passives are translated by Thai active constructions, but in only a tiny number of cases (4 out of 252, or slightly less than 1.6%) are agents that are only implicit in the source made explicit in the target. This means that in general, the conversion of passives into actives in our parallel corpus does not generally coincide with the kind of quantitative explicitation called ‘addition’ by Faber and Hjort-Pederson (2013: 44; and see Chapter 2).

In the 63% of cases where English passives are translated by Thai passives, we found that this is mostly achieved using the covert passive, which, again, does not involve explicitation. In fact, uses of covert passive outnumber uses of overt passive almost two to one. In the 54 cases in our data (or 21% of cases) where an English passive *is* translated using a Thai overt passive, however, that is using ถูก /thuk/ or ได้รับ /dai+/rup/, we tentatively claim that explicitation can be observed. We are tentative in this claim as it is difficult to say whether there is a self-evident increase in the amount of textual material in the target language vis-à-vis the source language, as required in the additive kind of explicitation recognized by Faber and Hjort-Pederson (ibid.) and discussed above: ถูก /thuk/ or ได้รับ /dai+/rup/ coincide with morphological markers of passive in English after all – namely forms of the auxiliary verb BE and past participles. Having said that, the particular markers that Thai uses often appear to reinforce or repeat meanings already implicit in the lexical verbs with which they collocate, *as well as* signalling passive voice, and on this basis their use might be seen as additive and explicitating. Given the fairly marginal use of these passive markers in translations of English passives into Thai, however, along with the extremely infrequent addition of subject-agents in cases where English passives are translated by Thai active clauses, we can say that the translation of passive voice from English into Thai provides only very weak evidence to support the hypothesis that explicitation is a feature of translation in the text type, language pair and direction under consideration.

### ***Textual Fit***

There is no straightforward linguistic counterpart for *ถูก* /thuk/ and *ได้รับ* /dai+/rup/ in English, so if we are analyzing translations from English into Thai, we consider *ถูก* /thuk/ or *ได้รับ* /dai+/rup/ as a unique item, which we in turn expect to be underrepresented in such translations (compared to native Thai texts). Specifically, if we were to find that the above-mentioned passive markers are used in translation into Thai less frequently than in comparable original texts in Thai, we would have evidence to support the unique items hypothesis. If, however, we see more instances of passive voice in translated Thai than in non-translated Thai, perhaps because the source texts ‘shine through’ (Teich 2003), then we may expect to see more frequent use of passive markers in translated text than in non-translated Thai text, thus contradicting the unique items hypothesis.

Our monolingual comparable data provides strong evidence *against* Tirkkonen-Condit’s (2002) unique items hypothesis as these two Thai passive markers are over-represented rather than under-represented in our translated data compared to comparable target-language data. The relative frequency of the *ถูก* /thuk/ passive in our Thai legal translations is almost twice that of our non-translated texts, and the relative frequency of the *ได้รับ* /dai+/rup/ passive in the translations is around 40% higher than that of the monolingual Thai legal corpus.

This finding could be due to the source language ‘shining through’, to use Teich’s (2003) expression. That is, the over-representation of passive markers in legislation translated into Thai could simply be due to the over-representation of passive voice in general in such translations, brought about by the influence of their passive-laden source texts. Pursuing this idea would require a larger-scale study of passives of all kinds (that is, not just of overt passives) in our Thai monolingual comparable corpus. It would also require us to investigate whether there are instances of passive voice in our Thai translations that do not coincide with passive voice in English.

### **Conclusions**

Drawing on previous corpus-based studies concerned with translation features, this chapter investigates whether there is evidence of explicitation (Becher 2010a) and source-language shining through (Teich 2003) in our translated data and whether

‘unique items’ (Tirkkonen-Condit 2002) are underrepresented. We pursue these questions through a study of the translation of passive voice into Thai, and focus in particular on the Thai overt passive markers ถูก /thuk/, โดน /don/ and ได้รับ /dai/+/rup/, which we consider unique items in Thai.

The study finds that, despite some writers’ antipathy to passive voice in Thai, most English passives are translated into Thai using passive voice, and that in those instances where active voice is used in the translation, there is rarely any explicitation involved, as explicit agents are rarely added in Thai. We do not find any evidence to support the hypothesis that unique items are underrepresented in translation. On the contrary, the unique items studied appear to be overrepresented in translation into Thai.<sup>14</sup>

Although this is considered first exploration of how passive voice is translated into Thai, and we have provided more empirical evidence on this topic than previous discussions, which were not based on corpora. However, our study leaves many questions unanswered, such as how the use of covert passive differs between translated and non-translated Thai. A study of this question (alongside further studies of overt passive) would doubtless make a useful contribution to the characterization of Thai national legislation as a legal-genre, as well as the description of translated legislation in Thai. It, however, remains beyond the scope of this thesis. This thesis also investigates unique items and explicitation through very narrow windows, and assumes comparability of normative non-translated legislative texts with merely informative translations. Having already created the corpus resources used in this study however, we are in a position to widen our investigation to consider other manifestations of passive voice as well as related linguistic features in translation from English to Thai, and to continue refining our analyses.

---

<sup>14</sup> The result could call into question the very status of /thuk/ and /dai/+/rup/ as unique items but that would be a logical error as it would suggest that unique items are unique because they are underrepresented in Thai translations; this is not the case. Such items in Tirkkonen-Condit (2002, 2004) are considered unique items because they do not have straightforward counterparts in the source language; their representation in the target language is a different issue.

## Chapter 5: Deontic Modality

### Introduction

In linguistic theory, a sentence used in communication is usually seen as consisting of a ‘proposition’, describing a state of affairs, and a ‘modality’, expressing concerns that the speaker or writer has towards that proposition (Iwasaki and Ingkaphirom 2005: 133). Two broad types of modality are commonly recognized: ‘epistemic modality’ is described by van Fintel (2006: 20) as concerning “what is possible or necessary given what is known and what the available evidence is”; while ‘deontic modality’ concerns “what is possible, necessary, permissible, or obligatory, given a body of law or a set of moral principles or the like” (van Fintel 2006: 20).<sup>15</sup> It is perhaps unsurprising then that the expression of deontic modality is “a key generic feature” of legislation (Biel 2014: 158). As Kurzon (1986: 15-16, quoted in Biel 2014: 158) puts it, legislative sentences are, after all, “speech acts with the illocutionary forces of permission (may), ordering (shall) or prohibition (shall not).” Despite the obvious importance of deontic modality in legislation, to date few studies have been conducted on deontic modals in legislative *translation*, and those studies that do exist appear to be based solely on European languages (e.g., Knežević and Brdar 2011; Biel 2014). This study breaks new ground by investigating deontic modality in our purpose-built parallel corpus of legislative texts translated from English into Thai. It also draws on our monolingual corpus of legislative texts originally written in Thai to enable comparisons between translated and non-translated texts in that language. After comparing the distribution of individual modal verbs in the English and Thai texts in the parallel corpus, on the one hand, and in the Thai translations and non-translated texts, on the other, to see how deontic modality is normally expressed in translation into Thai, we proceed to focus on ‘modal strength’ in translation, drawing on previous treatments that recognize strong and weak deontic obligation and permission (e.g. Palmer 1986).

This chapter is structured as follows: the next section addresses deontic modality. It provides an overview of previous treatments of deontic modality in Thai, and in translation generally, and further explains the notion of modal strength. Our research questions are then set out and our approach to data extraction, sampling and analysis is

---

<sup>15</sup> Palmer (1986: 102) adds a third category, namely ‘dynamic modality’, which is concerned with the ability and disposition of a participant.

explained, before detailed results are presented. The chapter concludes with some observations on the limitations of the research, and ideas for future work.

## **Deontic Modality in English and Thai**

The linguistic study of deontic modality traditionally focuses on modal verbs (Aijmer 2016: 496), also known as ‘modal auxiliaries’ or just ‘modals’. English modal verbs are often grouped into (a) ‘core modal verbs’ namely: can, could, may, might, will, shall, would, should and must,<sup>16</sup> and (b) ‘semi-modal verbs’, e.g. dare, need, ought to and used to (Carter and McCarthy 2006: 638). In English, deontic modal auxiliaries include verbs that express deontic obligation (sometimes referred to as ‘deontic necessity’), for example, ‘shall’, ‘must’ and ‘will’, and verbs that express deontic permission (sometimes referred to as ‘deontic possibility’), for example ‘may’ and ‘can’.<sup>17</sup> The corpus-based investigation of deontic modality is complicated by the fact that, in many languages, including English and Thai, a single form can express more than one type of modality (for example, the modal verb ‘must’ can be used either epistemically or deontically). While it is often easy to distinguish, for example, between epistemic and deontic uses based on formal criteria, in some instances the polyfunctionality of modal forms can nonetheless lead to interpretative ambiguity (Palmer 1986: 19).

Like English modality, Thai modality is traditionally divided into epistemic and deontic types (Iwasaki and Ingkaphirom 2005: 133), and can also be realized by modal verbs and other modal expressions. With specific regard to deontic modality, Iwasaki and Ingkaphirom (2005: 133-137) recognize four modal auxiliaries expressing obligation, namely: ต้อง /tong/ (must), จำเป็น /cham/+/pen/ (be necessary), ควร /khuān/ (should) and น่า /na/ (should).<sup>18</sup> Although other modal auxiliaries can be used to express deontic modality—as we will see below—these other modals are not listed by Iwasaki and Ingkaphirom (2005), perhaps because of their polyfunctionality.

---

<sup>16</sup> While there is debate in the literature over the status of ‘will’ as a modal (see, for example, Salkie 2010), the use of this verb in legislative texts to express obligation (much like ‘shall’ does) encourages us to view ‘will’ as deontic modal (see Table 5.5).

<sup>17</sup> Other parts of speech may also be used to express modality, but as it transpires that there are no relevant deontic examples in our corpus, we do not elaborate on this point here.

<sup>18</sup> Thai lexemes are presented here and elsewhere with an accompanying phonetic transcription (between two forward slashes) and an English-language gloss (in parentheses).



The co-occurrence of modal verbs is acceptable in Thai. The ‘serial verb’ จะต้อง /cha+/tong/, for example, does not have a straightforward counterpart in English, but it can be glossed as ‘will’ + ‘shall’, ‘must’ or ‘have to’.<sup>19</sup> (For the sake of brevity, we usually gloss this serial verb using ‘will+must’.) Knežević and Brdar (2011: 120) note a similar phenomenon in Croatian, which also allows the concatenation of modals in structures like *Moraš moći učiti* ‘You must can learn’. Even though similar structures exist in other languages like Croatian, the fact that English does not allow serial verbs and so has no structure corresponding to จะต้อง /cha+/tong/ means that from the point of view of English-to-Thai translation จะต้อง /cha+/tong/ is a ‘unique item’ as first proposed by Tirkkonen-Condit (2002), and discussed in Chapter 2.

### Modal Strength

As mentioned earlier, linguists often refer to ‘modal strength’, and view modal forms as capable of expressing ‘strong’ or ‘weak’ necessity or possibility (Narrog 2016), or something in between these poles. Some linguists refer to ‘degrees of modality’ (see Narrog 2009: 64-65) and organize modal expressions into scales along which expressions increase or decrease in degree. Halliday and Matthiessen (2004: 116, 148-149), for example divide deontic modal operators into ‘high’ (e.g. ‘must’), ‘median’ (e.g. ‘will’, ‘should’) and ‘low’ (e.g. ‘can’, ‘may’) categories. Huddleston (2002: 175-177) distinguishes between ‘necessity’ (e.g. ‘must’, ‘need’), ‘possibility’ (‘may’, ‘can’) and ‘medium modality’ (‘ought to’), while Knežević and Brdar (2011) grade deontic necessity according to ‘degree of commitment’ and use the terms ‘obligation’ for strong obligation, ‘necessity’ for weaker obligation, and ‘advisability’ where there is an implication that an obligation may not be fulfilled, among others.<sup>20</sup> According to Narrog (2016: 100), the expression of different levels of modal strength appears to be limited to the area of necessity. Thus, there are numerous descriptions of languages using labels such as strong and weak obligation but strong and weak permission are generally not found (ibid.). We likewise take the view that obligation modals can be graded according to their modal strength, but it makes little sense to compare the strength of permission modals. Having said that, between category-comparisons are valid (as in Knežević and

---

<sup>19</sup> A serial verb is defined by Iwasaki and Ingkaphirom (2005: 231) as “a grammatical structure in which two or more verbs or verb phrases appear together without a marker of coordination or subordination”.

<sup>20</sup> Narrog (2009:64-65) gives a useful summary of approaches to modal strength or degree of modality.

Brdar 2011), so it is possible to say that any given obligation modal is stronger than any given permission modal, even if one permission modal cannot be stronger than another permission modal.

Thai deontic modal verbs also express different degrees of strength, with Iwasaki and Ingkaphirom (2005: 133-137) claiming that ต้อง /tong/ (must) expresses the strongest obligation, followed by จำเป็น /cham/+/pen/ (be necessary), ควร /khuan/ (should) and น่า /na/ (should).<sup>21</sup>

### **Modality and Modal Strength in Translation**

Deontic modality in legislative translation has been studied by Biel (2014), who compares EU legislation translated from English into Polish with original Polish legislative texts. Biel's research design involves both a 'parallel' corpus of English legislative source texts and their translations into Polish, and a 'comparable' corpus of legislation originally written in Polish, a design that is emulated in our work. Biel is most interested in establishing the 'textual fit' of the translations, where 'textual fit' is a target-pole oriented measure defined as the 'linguistic distance between translations and non-translations of a comparable genre' (Biel 2014: 287). Textual fit complements 'equivalence', defined simply as the relation of translations to their source texts, in that they are both facets that are critical for the quality of translation (ibid.). As part of her study, Biel investigates the distributions of various modal forms in her corpora, and the possible reasons for differences between translations and non-translations. She finds that obligation modals (verbs) are 'strongly overrepresented' in translated Polish compared to original Polish, while deontic phraseological patterns are 'strongly underrepresented' (Biel 2014: 289). She concludes that "modals are strongly affected by interference and translators rarely strive to overcome it and conform to the conventions of national legislation" (2014: 169). Biel also finds that "Differences in the distribution of deontic modals across the corpora indicate that modality is very sensitive to goals of sub-genres" (ibid.). She does not investigate modal strength as such.

---

<sup>21</sup> It should be mentioned here that the range of modal verbs imposing obligations and indicating permission in Thai is smaller than in English, and a single Thai modal can thus be used to gloss more than one English modal. For example, ต้อง /tong/ is the Thai gloss of 'shall', 'must' and 'have to'. ควร /khuan/ is the Thai gloss of 'should' and 'ought to'. Conversely, more than one English modal can gloss a given Thai modal.

Knežević and Brdar (2011) and Martikainen (2016) are among the few sources that investigate modal strength in translated texts, although only the former are concerned specifically with deontic modality. They investigate deontic modal verbs in a 23,000 word corpus consisting of a piece of Croatian legislation, *Plan prihvata broda u nevolji*, and its translation into English as ‘Plan for the Acceptance of a Ship in Distress’ and observe shifts in modal strength in the translation. Most shifts (7 out of 113 instances) relate to an intensification (Wilcox and Shaffer, 2008: 229) from ‘necessity’ in Croatian to ‘obligation’ in English. Although shifts in the opposite direction are also observed, they are fewer in number at two (Knežević and Brdar 2011: 135-136). While the amount of data that Knežević and Brdar (2011) examine is limited, their results are still interesting because they show that shifts in modal strength, while rare, are attested in legislative translation. The authors conclude that the reason for such shifts lies in “the lack of clear and unitary syntactic rules in Croatian” with the use of two particular modal verbs in Croatian described as “generally widespread” but “arbitrary” (2011: 142). Knežević and Brdar (2011) also stress the importance of context in the interpretation of the deontic meaning of particular verbs.

Martikainen (2016) investigates translation errors and biases in human and machine translations of medical abstracts from English into French. She homes in on markers of epistemic modality, and thus is concerned with scales of certainty rather than obligation and permission. Her research is nevertheless of interest here as she shows how some English modal markers are translated into more affirmative, and thus ‘stronger’ French counterparts. More specifically “the modal auxiliary ‘should’ is frequently translated into French by the indicative form of the verb ‘*devoir*’ in the present or the future tense (equivalent of ‘must’), instead of the corresponding conditional form” (Martikainen 2016: 167). In most cases, she argues, the translations do not lead to bias, as these more affirmative modal markers are typically used in the target context and are likely to be expected and correctly interpreted by the readers (ibid.); that said, in some cases the more affirmative translations can influence the readers’ interpretation of the level of certainty of the authors (ibid.).

The present study is the first attempt, to our knowledge, to investigate the translation of modality and degrees of modal strength in translation between English and Thai legislative texts.

## Research Questions

In this study we are interested in both the relationship between source texts and translations ('equivalence' relations as defined by Biel 2014: 287), and between target texts and other texts in the same language (Biel's 2014 'textual fit'). We set out to answer two main questions:

1. How are instances of deontic modality translated from English into Thai in legislative texts?
2. Are there differences between how deontic modality is typically expressed in legislation translated into Thai and how it is expressed in legislation originally written in Thai?

## Research Procedure

Our investigation begins with the identification of modal verbs in the English side of our parallel corpus, more or less following the methodology used by Biel (2014). Using Sketch Engine's Wordlist function, we retrieve a frequency-ranked lemmatized list of all words (types) in the corpus that have been assigned the part-of-speech tag 'MD' (for 'modal'). Given the polyfunctionality of modal verbs addressed earlier, we check (using the software's Concordance view) a random sample of the hits to see if any of them can be interpreted epistemically. After verifying that is not the case, we start examining the frequency of deontic modals in the English sub-corpus. For the sake of completeness we also search for other forms of modality in our corpus: again using the Wordlist function in Sketch Engine, we generate a list of candidate modal adverbs/adjuncts and modal adjectives as identified by Palmer (1986). Inspection of the results from this search shows that there are a small number of epistemic modal adverbs and adjectives in our corpus, but there are no deontic uses of these word classes. Our analysis is subsequently restricted to deontic modality as realized by modal verbs.

Having established the distribution of deontic modal verbs in the English side of our parallel corpus, we then compare the English lines containing each of these verbs with their Thai translations. We are interested here in how English modals are translated generally, but also in whether there are shifts in modal strength as identified in previous studies. In a final step, we compare our findings from the parallel corpus with data from our monolingual Thai corpus to see if the modal verbs used in the Thai translations are

overrepresented or underrepresented when compared with the Thai monolingual corpus. On the Thai side of the parallel corpus and in the Thai monolingual corpus, we cannot use the Wordlist function to find the frequency of modal verbs due to limitations of the word tokenizer and the lack of a part-of-speech tagger for Thai. We thus conduct direct searches for the Thai deontic modals which have been found as translations of English modal verbs in the parallel corpus. These searches are conducted using Sketch Engine's Concordance tool.

A final word here is said about our sampling strategy: as already indicated, given the polyfunctional nature of most modal verbs, it is essential to check the context in which modal forms occur, to verify that they are, indeed, deontic uses of the forms in question. This requires the analyst to inspect concordances from the corpus, which can be a time-consuming process. In cases where there are too many instances of a form to make manual inspection of all of them possible, we generate a random sample of instances (using the Random Sample function in Sketch Engine), and base our judgments on this sample. Our random samples consist of either 100 instances of each form, or 10% of the total number of occurrences for that form (for extremely frequent forms), as illustrated respectively by 'may' and 'shall' in Table 5.2.

## **Results and Analysis**

In this section we present results first from the English side of our parallel corpus, then from the Thai translations, and finally we compare our Thai translations with original Thai in our monolingual corpus.

### ***Deontic modal verbs in the English source texts***

The lemmatized frequency-ranked wordlist output by Sketch Engine indicates that there are 4,337 occurrences (tokens) of ten modal verbs (types) in our English source texts. Their distribution is summarized in Table 5.1.

<b>Modal verbs</b>	<b>Number of occurrences</b>	<b>Percentage</b>
1. shall	3,202	74
2. may	745	17
3. should	131	3
4. will	102	2.5
5. can	61	1.4
6. would	52	1.2
7. must	24	0.55
8. could	11	0.25
9. might	6	0.13
10. ought to	3	0.07
Total	4,337	100

**Table 5.1: Frequency of modal verbs in the English source texts**

The most striking result in Table 5.1 is the very high frequency of ‘shall’. There are 3,202 instances of ‘shall’ accounting for 74% of all instances of modal verbs in the corpus. Using the Wordlist function to find the overall frequency ranking of ‘shall’ in our English source texts, we find that ‘shall’ comes in sixth place, after ‘the’, ‘of’, ‘and’, ‘to’ and ‘in’. The frequency of ‘shall’ is thus very high in this small English corpus.

The second most frequent modal verb in our English source texts is the deontic permission verb ‘may’, accounting for 17% of all modal verbs. The relative frequencies of the remaining modal verbs in our corpus (‘should’, ‘will’, ‘can’, ‘would’, ‘must’, ‘could’, ‘might’ and ‘ought to’) are at 3% or lower.

We further investigate only the five most frequent deontic modal verbs, dividing them into two groups—deontic obligation (‘shall’, ‘should’ and ‘will’) and deontic permission (‘may’ and ‘can’). With the exception of ‘can’, which tends to be used

epistemically or dynamically, the vast majority of the forms in question function as instances of deontic modality in context, as illustrated in Table 5.2:

<b>Modal verb</b>	<b>Number of occurrences</b>	<b>(Random) sample size</b>	<b>Functioning as deontic modal verbs</b>
shall	3202	320	320
should	131	131	126
will	102	102	101
may	745	100	100
can	61	61	12

**Table 5.2: Deontic obligation and permission modal verbs**

*Thai translations of English deontic modal verbs*

Table 5.3 shows the overall distribution of Thai translations of the modal ‘shall’.

<b>Thai translations of ‘shall’</b>	<b>English gloss</b>	<b>Number of occurrences</b>	<b>Percentage</b>
simple declarative	subject+verb	93	29
จะ /cha/	will	93	29
จะต้อง /cha+/tong/	will+must	45	14
Thai causative starting with ให้ /hai/ (hai structure)	hai+subject+verb hai+verb+complement	35	11
ต้อง /tong/	shall/must/have to	35	11
ควร /khuan/	should	13	4
อาจ /aat/	may	2	0.6
lines not translated	-	4	1.4
total		320	100

**Table 5.3: Thai translations of deontic modal ‘shall’**

As Table 5.3 indicates, based on our random selection of 320 concordance lines, we find that ‘shall’ is translated into Thai as follows: 29% of instances are translated into a Thai simple declarative form. A further 29% are translated using *จะ* /cha/ (will). Translation of ‘shall’ into the previously mentioned serial modal *จะต้อง* /cha+/tong/ (will+must) accounts for 14%, while 11% of instances of ‘shall’ are translated by a Thai causative form. The closest Thai equivalent of ‘shall’,<sup>22</sup> *ต้อง* /tong/, appears in only 11% of cases. The translators also employ weaker deontic modals, namely *ควร* /khuan/ (should) and *อาจ* /aat/ (may), accounting for 4% and 0.6% of cases respectively.

Example 5.1 illustrates the use of different Thai verbs to translate ‘shall’.

### *Example (5.1)*

English source text:

This Charter *shall* be signed by all ASEAN Member States. This Charter *shall* be subject to ratification by all ASEAN Member States in accordance with their respective internal procedures. Instruments of ratification *shall* be deposited with the Secretary-General of ASEAN who *shall* promptly notify all Member States of each deposit. This Charter *shall* enter into force on the thirtieth day following the date of deposit of the tenth instrument of ratification with the Secretary-General of ASEAN.

Thai translation:

กฎบัตรนี้*จะต้อง*ได้รับการลงนามโดยรัฐสมาชิกอาเซียนทั้งหมด กฎบัตรนี้*จะอยู่*ได้บังคับของการให้สัตยาบันจากรัฐสมาชิกอาเซียนทุกรัฐตามกระบวนการภายในของแต่ละรัฐ สัตยาบันสาร*จะต้อง*เก็บรักษาไว้กับเลขาธิการอาเซียน ซึ่งจะแจ้งให้รัฐสมาชิกทุกรัฐทราบถึงการส่งมอบสัตยาบันสารแต่ละฉบับโดยพลัน กฎบัตรนี้*จะมีผล*ใช้บังคับในวันที่สามสิบหลังจากวันที่มีการส่งมอบสัตยาบันสารฉบับที่สิบให้แก่เลขาธิการอาเซียน

---

<sup>22</sup> By ‘closest equivalent’ here, we mean the form that we would expect to be used as an equivalent, on the basis of Thai native-speaker intuition.



Back translation (gloss):

charter this /*cha*/+/*tong*/ (will+ must) /*dai*/+/*rup*/ signature by state member  
 asean all. charter this /*cha*/ (will) subject to ratification from state member  
 asean in accordance with procedure internal of each state. Instrument  
 /*cha*/+/*tong*/ (will+ must) deposit with secretary-general asean who /*cha*/ (will)  
 notify state member all to know about deposit of instrument each promptly.  
 charter this /*cha*/ (will) enter into force on thirtieth day after date of deposit of  
 instrument tenth to secretary-general asean

There are 126 instances of deontic ‘should’ in the English texts and we investigate all concordance lines.

Thai translation of ‘should’	English gloss	Number of occurrences	percentage
ควร / <i>khuan</i> /	should	106	84.1
ต้อง / <i>tong</i> /	shall/must/have to	5	4
จะ / <i>cha</i> /	will	5	4
จะต้อง / <i>cha</i> /+/ <i>tong</i> /	will+ must	4	3
simple declarative	subject+verb	3	2.4
อาจ / <i>aat</i> /	may	1	0.8
lines not translated	-	2	1.6
total		126	100

**Table 5.4: Thai translations of deontic modal ‘should’**

Table 5.4 shows that the majority of instances of ‘should’ (approximately 84% of cases) are translated by the Thai modal ควร /*khuan*/, as expected. In 11% of cases ‘should’ is translated using stronger modals however, namely ต้อง /*tong*/ (must) (4%), จะ /*cha*/ (will) (4%), จะต้อง /*cha*/+/*tong*/ (will+ must) (3%), and or simple declarative (~2%), and in one case, it is translated into the deontic permission modal, อาจ /*aat*/ (may).

Example 5.2 illustrates the use of ต้อง /*tong*/ (must) to translate ‘should’.

### Example (5.2)

English source text:

Every citizen has the right to vote in periodic and genuine elections, which *should* be by universal and equal suffrage and by secret ballot, guaranteeing the free expression of the will of the electors, in accordance with national law.

Thai translation:

พลเมืองทุกคนมีสิทธิลงคะแนนในการเลือกตั้งตามกำหนดเวลาและอย่างแท้จริง ซึ่งต้องเป็นการเลือกตั้งอย่างทั่วถึงและ  
เสมอภาค และต้องเป็นการลงคะแนนลับ โดยได้รับการประกันการแสดงออกถึงเจตจำนงของผู้มีสิทธิเลือกตั้งอย่างเสรีตาม  
กฎหมายของแต่ละประเทศ

Back translation (gloss):

citizen every has right to vote in election periodic and genuine which /tong/  
(must) be election universal and equal and must be ballot secret receive  
guarantee expression free of will of elector in accordance with law of each  
country.

There are 101 occurrences of ‘will’ functioning as deontic modals in the English sub-corpus. Their Thai translations are shown in Table 5.5.

Thai translation of ‘will’	English gloss	Number of occurrences	percentage
จะ /cha/	will	67	66.3
simple declarative	subject+ verb	27	26.7
ต้อง /tong/	shall/must/have to	4	4
ควร /khuan/	should	2	2
จะต้อง /cha/+/tong/	will+ must	1	1
total		101	100

Table 5.5: Thai translations of deontic modal ‘will’

66% of instances of ‘will’ are translated as จะ /cha/, as expected. Over a quarter (27%) of instances of ‘will’ are translated into simple declarative without a modal verb. The remainder are translated into the stronger modals ต้อง /tong/ (must) (4%) (as illustrated in Example 5.3), ควร /khuan/ (should) (2%), จะต้อง /cha+/tong/ (will+must) (1%).

**Example (5.3)**

English source text:

These carriers *will* be jointly and severally liable to the passenger or to the consignor or consignee.

Thai translation:

ผู้ขนส่งทุกทอดที่กล่าวมานั้นย่อม**ต้อง**รับผิดชอบร่วมกันและแทนกันต่อคนโดยสาร หรือต่อผู้ตราส่งหรือผู้รับตราส่ง

Back translation (gloss):

carriers mentioned */tong/* (must) be liable jointly and severally to passenger or to consignor or consignee.

Table 5.6 shows Thai translations for the deontic permission verb ‘may’.

Thai translation of ‘may’	English equivalent	Number of occurrences	percentage
อาจ /aat/	may	64	64
simple declarative	subject+verb	13	13
สามารถ /sa-mart/	can	10	10
ต้อง /tong/	shall/must/have to	3	3
จะ /cha/	will	5	5
ควร /khuan/	should	1	1
lines not translated		4	4
total		100	100

**Table 5.6: Thai translations of deontic modal ‘may’**

‘May’ occurs 745 times in the English sub-corpus, so we investigate a random sample of 100 concordance lines. 64% of instances are translated as the Thai modal อาจ /aat/, as expected. 13% of instances use simple declaratives with no modal verb to translate sentences with ‘may’. 10% of instances are translated by another deontic permission modal, namely สามารถ /sa-mart/ (can), while 9% of instances are translated using deontic obligation verbs in Thai (see Example 5.4).

**Example (5.4)**

English source text:

The provisions of paragraphs 1 and 2 of the present article *may* not be interpreted as adversely affecting any right to stay and work otherwise granted to such family members by the legislation of the State of employment or by bilateral and multilateral treaties applicable to that state.

Thai translation:

บทบัญญัติในวรรค 1 และ 2 ของมาตรานี้ *ต้อง*ไม่มีการนำไปตีความหมาย อันมีผลเป็นอุปสรรคต่อการพำนักและการทำงาน นอกจากสมาชิกครอบครัวนั้นจะได้รับอนุญาตตามกฎหมายของรัฐที่มีการจ้างงาน หรือโดยสนธิสัญญาทวิภาคีและพหุภาคีที่ใช้กับรัฐนั้น

Back translation (gloss):

Provision in paragraph 1 and 2 of article this */tong/* (must) not interpretation which result effect adverse to right to stay and work except member family such receive permission by law of state of employment or by treaties bilateral and multilateral application to state that.

There are twelve instances of deontic ‘can’ in the English sub-corpus (Table 5.7). The most frequent translation is, as expected, สามารถ /sa-mart/, accounting for ten instances, while one instance is translated as the Thai deontic obligation จะ /cha/ (will) (Example 5.5) and another one by the Thai causative ‘hai+verb+complement’.

Thai translation of 'can'	English equivalent	Number of occurrences	Percentage
สามารถ /sa-mart/	can	10	84
จะ /cha/	will	1	8
Thai causative starting with ให้ /hai/ (hai structure)	hai+verb+complement	1	8
total		12	100

**Table 5.7: Thai translations of deontic modal 'can'**

**Example (5.5)**

English source text:

This penalty *can* only be carried out pursuant to a final judgement rendered by a competent court.

Thai translation:

การลงโทษเช่นนี้จะกระทำได้อีกแต่โดยคำพิพากษาถึงที่สุดของศาลที่มีอำนาจ

Back translation (gloss):

penalty such /*cha*/ (will) carry out according to judgement final of court authorized.

***Deontic Modals in Monolingual Thai and Thai Translation***

Table 5.8 compares the frequencies of selected deontic modal verbs in the Thai translations in our parallel corpus and in our comparable monolingual Thai corpus. The selected modal verbs are the six Thai modal auxiliaries that appear as translations of the English modals under investigation in the previous section, namely:

1. ต้อง /tong/ (an equivalent of 'shall', 'must' or 'have to')
2. จะต้อง /cha+/tong/ (an equivalent of 'will'+ 'must')
3. จะ /cha/ (an equivalent of 'will')
4. ควร /khuan/ (an equivalent of 'should' or 'ought to')

5. อาจ /aat/ (an equivalent of ‘may’)
6. สามารถ /sa-mart/ (an equivalent of ‘can’)

Thai modal verbs	Thai monolingual corpus (1,173,485 words)	Thai translation corpus (222,556 words)	Log-likelihood (LL) score
	Number of occurrences (frequency per million)	Number of occurrences (frequency per million)	
ต้อง /tong/ (shall/ must/ have to)	4,899 (4,174.74)	598 (2,686.96)	116.04
จะต้อง /cha+/tong/ (will+ must)	828 (698.77)	619 (2,781.32)	585.15
จะ /cha/ (will)	4630 (3,948.51)	1743 (7,831)	531
ควร /khuan/ (should)	151 (128.67)	274 (1,231.15)	505.63
อาจ /aat/ (may)	1,209 (1,030.26)	571 (2,565.64)	283.11
สามารถ /sa-mart/ (can)	704 (599.9)	172 (772.83)	8.42

**Table 5.8: Thai deontic modals in translation and monolingual sub-corpora**

From the figures in the Table 5.8, based on the basic descriptive statistics, we find that 5 out of 6 modals are overrepresented in the translations when compared with the Thai monolingual texts, while one is underrepresented. We test for statistical significance using the UCREL log-likelihood wizard, created by Paul Rayson. According to the UCREL log-likelihood (LL) test,<sup>23</sup> if the LL score is greater than 6.63, the probability of the result happening by chance is less than 1%, and if the LL score is 3.84 or more, the probability of it happening by chance is less than 5%, and we are 95% certain of the result. As such, the LL score must be above 3.84 for the difference to be significant. All differences reported in Table 5.8 reach statistical significance.

<sup>23</sup> The detail of the calculating method can be found in the website of the University of Lancaster regarding Statistics in Corpus Linguistics: <https://corpora.lancs.ac.uk/clmtp/2-stat.php>.

Of the overrepresented forms, *จะต้อง* /cha/+/tong/ (will+must) is almost four times more frequent in the Thai translations (2,781.32 occurrences per million words) than in the Thai monolingual corpus (698.77 per million words). Likewise, the frequency of *จะ* /cha/ (will) in translation (7,831 per million words) is about five times higher than that in original texts (3,948.51 per million words). The frequency of *ควร* /khuan/ (should) in Thai translation texts (1,231.15 per million words) is almost 10 times higher than that in the Thai monolingual texts (128.67 per million words). The relative frequency of the permission modals *อาจ* /aat/ (may) and *สามารถ* /sa-mart/ (can) is slightly higher than that in the monolingual texts.

There is only one modal verb, namely *ต้อง* /tong/ (shall/must/have to), that is underrepresented in translation. Its frequency in the Thai translation is slightly lower than that in the Thai monolingual texts.

## Discussion

### *Source-Language Modals*

Our first finding of note, relates to the high frequency of occurrence of ‘shall’ in our English source texts, accounting for 74% of all instances of modal verbs. Given the use of ‘shall’ to impose strong obligations, its high frequency in legislative texts may not be surprising. It is worth remembering, however, that the Plain Language Movement that took hold in the major English-speaking countries in the 1970s, aimed, among other things, to rid legal English of the word ‘shall’ (Williams, 2009, 199-208). The modal ‘shall’ has thus been eliminated from a number of prescriptive legal texts particularly in the major English-speaking countries of the southern hemisphere—Australia, New Zealand and South Africa; nevertheless, the major international organizations in the northern hemisphere such as the United Nations, the International Labour Organisation or the European Union still prefer to use the archaic word ‘shall’ in their legislative texts (Williams 2009, 200). The fact that our English sub-corpus contains only international treaties which have had legal force since the 1950s, and the majority of which were drafted by the United Nations, might explain why there is such a high frequency of deontic modal ‘shall’ in our English corpus. Similar results are reported by Biel (2014: 159), who finds that ‘shall’ accounts for two thirds of obligation modals found in the English-language legislative texts in her study, which emanate from the

European Union. Our finding that ‘may’, the second most common deontic modal in our English texts, accounts for 17% of deontic modals in our corpus is also broadly in line with Biel (2014: 166-169), who finds that permission modals are significantly less frequent than obligation modals in her English legislative corpus.

### ***Relations between Source and Target Texts***

Turning now to the translation into Thai of English modal verbs, and concentrating firstly on relations between source and target texts (Biel’s (2014) ‘equivalence’ relations), we can make the general claim that almost half (40%) of all instances of the most common deontic modal in our corpus, ‘shall’, are translated in a way that departs from formal equivalence (or ‘formal correspondence’ in Catford’s (1965) terminology). In other words, the translations do not use modal verbs to translate modal verbs. Rather, these instances of ‘shall’ are translated into Thai non-modal structure, either in declarative form (29%) or causative form with the word *ไห้* /hai/ (11%).

As already indicated, the expression of deontic modality is not limited to modal verbs, and according to Williams (2009), under the Plain Language Movement, the use of present simple is encouraged as a replacement of ‘shall’ to lay down general principles in legal texts. The translation of sentences containing the English modal ‘shall’ into Thai non-modal sentences also seems appropriate, given that the obligation can be inferred from the genre itself. We note however, that such translations could constitute cases of implicature, where ‘implicature’ refers to the non-verbalization in a target text of information that was verbalized in the source text, but that the target-language addressee might be able to infer from the context or other inferential source anyway (see Becher 2010a, 2010b, 2011). It is difficult to make generalizations about the modal strength of (non-modal) simple declarative sentences in the Thai translations however, especially because they are also used to translate instances of the ‘weaker’ deontic obligation verbs ‘will’ (27 instances, or 26.7%) and ‘should’ (3 instances, or 2.4%), and even the deontic permission verb ‘may’ (13 instances, or 13%). One possible interpretation is that if simple declaratives are a common way of translating ‘shall’, then their slightly less common use to translate ‘will’ and infrequent use to translate ‘should’, may serve to elevate the obligation in the latter two cases.

The *ไห้* /hai/ causative form has already been encountered in Chapter 4. To recap, the placing of *ไห้* /hai/ at the beginning of a sentence creates a causative construction



(Iwasaki 1998) in which one entity, the ‘trigger’ (ibid.), causes, enables or obliges another entity, the ‘instigator’ (ibid.) to carry out some action. The *ให้* /hai/ causative can express different levels of obligation, from weak to strong, depending on the context, but we can argue that such causatives used in Thai legal texts are certainly not weak. This is because we can infer from the context that the trigger that demands the instigator to act is the law (treaty) itself. The translators’ use of a *ให้* /hai/ causative instead of a modal verb such as *ต้อง* /tong/ (the closest equivalent to ‘shall’) suggests they consider that the use of the causative form elevates the degree of obligation in the same way as the modal verb ‘shall’ does. The Thai causative structure as used in translated legislative texts is also performative and somewhat imperative because it indicates that a command is being issued. According to Palmer (1986, 108) the imperative is not necessarily stronger or weaker than the strong modal verbs, however, because it is presented only as a deontic proposition and it is the hearer who has to judge the force of his or her obligation to act from the circumstances. This makes it difficult to compare directly the modal strength of imperatives and modal verbs, but we might argue, as Becher (2011: 43-44) does, that the illocutionary force of imperatives and certain deontic modals is more or less the same.<sup>24</sup>

In a further 93 instances or 29% of cases, the modal ‘shall’ is translated by the Thai modal *จะ* /cha/ (will). The word *จะ* /cha/ in Thai is often translated as ‘will’ in English and behaves similarly to ‘will’ both syntactically and semantically (Srioutai 2004, 100). Similarly to ‘will’, *จะ* /cha/ (will) can stand for the future tense, but it is also compatible with expressions of different types of modality, including deontic necessity and deontic possibility, as argued by Srioutai (2004). This claim is relevant to the use of *จะ* /cha/ (will) in our parallel corpus, which involves deontic necessity in the future.<sup>25</sup> The use of *จะ* /cha/ (will) as a deontic modal does not necessarily make the obligation less intense.

Another 45 instances (14%) of ‘shall’ are translated into Thai by *จะต้อง* /cha+/tong/ (will+must). As previously mentioned, *จะต้อง* /cha+/tong/ is an example of a serial verb

---

<sup>24</sup> Becher (ibid.) also argues that imperatives are more explicit than modals such as ‘must’, given the polyfunctionality of such modals, as referred to earlier. He does not take into account however, the fact that ‘must’, while ambiguous in isolation, is often not at all ambiguous in context, and is certainly not ambiguous in the example he himself provides.

<sup>25</sup> Palmer (1998, 97) explains that deontic modal verbs are performative and always related to the future because only the future can be changed or affected as a result of them being expressed.

in Thai, and given that English lacks such structures, it can be considered a ‘unique item’ from the point of view of English to Thai translation. Tirkkonen-Condit (2004, 177-178) holds that such unique items “do not readily suggest themselves as translation equivalents, as there is no obvious linguistic stimulus for them in the source text.” On this basis, Tirkkonen-Condit (2004, 178) goes on to hypothesize that unique items are likely to be underrepresented in translated texts compared to non-translated texts in the same language. Our analysis so far has shown that *จะต้อง* /cha+/tong/ is definitely attested in translation from English into Thai, but not whether it is underrepresented in Thai. We return to this issue when we discuss ‘textual fit’ below.

Contrary to expectation, we find only 35 instances (11%) of *ต้อง* /tong/ (shall/must/have to) as translations of ‘shall’ in our corpus. Finally we note 13 instances (4%) where the weaker deontic obligation modal *ควร* /khuan/ (should) is used and two instances (0.6%) where the weaker deontic permission modal *อาจ* /aat/ (may) is used. ‘Shall’ is thus the only modal verb in our corpus where a weakening of modal strength occurs in translation, but at 4.6% of instances, this is a marginal phenomenon.

In the case of our other two obligation modals, we see that translations of ‘should’ and ‘will’ are somewhat more predictable than translations of ‘shall’ with 84% of instances of ‘should’ translated, as expected, by *ควร* /khuan/ and 66% of ‘will’ translated, again as expected, using *จะ* /cha/. Having said that, other translations of ‘should’ and ‘will’ elevate the deontic obligation in 11% and 7% of cases, respectively, and the use of simple declaratives in just over one quarter of instances also complicates the analysis for ‘will’, as described above.

Likewise, with verbs of deontic permission, the majority (64%) of instances of ‘may’ are translated as expected by *อาจ* /aat/, while 10% are translated by another deontic permission modal, namely *สามารถ* /sa-mart/ (can). Nine instances of ‘may’ (9%) are translated into expressions of deontic obligation in Thai, however, thus elevating the modal strength, and simple declaratives are also used to translate 13% of instances. Likewise, ten out of twelve instances of ‘can’ are translated, as expected, by *สามารถ* /sa-mart/ while the other two are translated by the Thai causative structure and the obligation modal *จะ* /cha/ (will), marking a shift from permission to obligation.

Summing up, just over 31% of instances of English deontic modals in our sample are

translated in a way that diverges syntactically from the source text, with simple declaratives accounting for 20.63% of these cases, and the Thai causative for a further 10.62%, but the semantic component of obligation remains intact in these instances, and the strength of the obligation seems undiminished, although comparisons of ‘modal strength’ are problematic here, for the reasons outlined above. As already indicated, we are on safer ground comparing the strength of obligation modals with each other, or of obligation modals with permission modals. Restricting ourselves to such comparisons, we note that modal strength is reduced in just under 2.28% of cases in our sample, and increased in just under 4.86%.

### *Textual Fit*

Our comparison between translated and the non-translated Thai found that five out of six Thai modal verbs investigated were overrepresented in the translations, echoing Biel’s (2014, 289) finding that obligation modals are overrepresented in Polish legislative translations.<sup>26</sup> Biel explains her finding in terms of interference from source texts. There is good reason to believe that our source texts are an influencing factor in the present study too. If we consider again the fact that there are 4,337 modal verbs in our 172,739 word English corpus (see Tables 5.1), then we would expect to see 25,107 modal verbs in a one-million-word corpus of legislative texts in English. Our Thai monolingual corpus (with 1,173,485 words), by way of contrast, contains 12,421 modal verbs (summed from the values in Table 5.8), or just under 10,585 per million words. Modal verbs are thus nearly 2.5 times more frequent in the English source texts in our corpus than they are in our non-translated Thai legislative texts.<sup>27</sup> Our translations into Thai (222,556 words) occupy an intermediary position, with 3,977 modals in all (again, summed from the values in Table 5.8), or 17,869 per million. Thus although Biel (2014, 289) claims that the translators represented in the parallel corpus she uses rarely strive to adjust to the norms of national legislation, it seems as if our Thai translators are actually steering a middle course between English and Thai distributions of modal verbs. Indeed, they end up close to halfway between the norms for the two languages.

---

<sup>26</sup> Again like Biel (2014), we find that this effect is much reduced in the case of deontic verbs of permission.

<sup>27</sup> Even if we allow for the fact that a small number of these English modals are epistemic (137 would be both an accurate and convenient estimate), we would still have 4,200 deontic modals in our English corpus, equivalent to 24,314 deontic modals per million words. We have no reason to believe that similar adjustments to the count for the Thai texts would change this ratio to any great extent.

We may speculate that the Thai translators are somewhat freer to move towards target-language norms precisely because, unlike the translators of the EU legislation addressed by Biel (2014), their translations are not ‘authentic’ in the sense described earlier (see also Dullion 2000; Garzone 2000).

Moving beyond these broad generalizations, it is still remarkable that the Thai modal verbs studied here are *so* much more frequent in translated than in non-translated Thai. We know already that sentences with modal verbs are in complementary distribution with other structures, for example simple declarative and causative sentences, in the Thai texts in our parallel corpus. This is, of course, also true of our monolingual Thai corpus, but we have not established the precise distribution for anything but instances of translations of selected English modal verbs. It is possible that the distribution in Thai translations is skewed by virtue of the source language shining through, to use Teich’s (2003) term, and that, for example, simple declaratives and causatives are more commonly used in non-translated Thai than in translated Thai, but we have yet to systematically pursue this line of inquiry. What is already clear however is that the serial modal verb จะต้อง /cha+/tong/ (will+must) is very much indicative of Thai legislative translation (with a log likelihood score of 585.15); rather than being underrepresented, as Tirkonnen-Condit’s (2002, 2004) unique items hypothesis would predict, it is overrepresented in Thai legislative translation. This finding might suggest that translated legislation as a genre is subject to textual-linguistic norms (Toury 1995: 58-59) that transcend language typological differences. The higher than expected frequency of other Thai modals in translation can also be partly explained by the fact that translations of ‘shall’, which is by far the most common modal in the English source texts, are so variable. Translations of ‘shall’ contribute, for example, to the high frequencies of จะ /cha/ (will) and even ควร /khuuan/ (should). Likewise, the frequency count for อาจ /aat/ (may) in the translations is elevated not just by the fact that this modal is used to translate nearly two thirds of instances of ‘may’, but also by its use to translate a range of other English modals including ‘shall’, ‘could’, ‘can’, ‘might’ as well as instances of the simple declaratives (with no modal verbs). A full description of deontic modality in Thai translation would have to take account of such many-to-many translations, and would need to approach the issue in both directions: from English to Thai, and from Thai to English, as well as making comparisons between translated and non-translated Thai. So far we have approached our study from just two of these three

angles. Finally, it is ironic that what we considered to be the most obvious translation of ‘shall’, namely *ต้อง* /tong/, accounts for only 11% of translations of ‘shall’, and is underrepresented in Thai legislative translation compared to Thai non-translated legislation. We can only speculate here that it is somehow displaced by competing modals that have become more indicative of translated legislation in Thai.

## **Conclusions**

This chapter examines how English deontic modal verbs are translated into Thai, and how translated Thai compares with non-translated Thai in the context of legislative translation. Our broad findings are that the translation of English modals into Thai shows more variability than we might have imagined, and that 31% of such modals are translated using resources other than modals in Thai. Given the use of simple declaratives to translate around 20% of forms, obligation can sometimes become implicit rather than explicit in Thai translation. Shifts in modal strength occur, but are relatively rare at just over 2% for decreases and just under 5% for increases. Modal verbs turn out to be overrepresented in translation into Thai compared with non-translated Thai, and we found no evidence to support the unique items hypothesis.

The study is limited by the fact that causatives and simple declarative structures that indicate obligation or permission have not yet been systematically studied, and that full explanations have yet to be arrived at for the distribution of modal forms in Thai. Such explanations would benefit from reversing the direction of the analysis, and proceeding from Thai into English. Finally, we have not considered the effects of sub-genres.

In future research we hope to pursue these ideas.

## **Chapter 6: Cohesive conjunction**

### **Introduction**

In this chapter, we shift focus to the linguistic phenomenon of cohesion, and more specifically to cohesive conjunction. The structure of this chapter is as follows. We start by defining the concept of cohesion and then briefly describe its realization in both English and Thai contexts. We then define the concept of cohesive conjunction as used in this research. After that, a brief description of our data processing and sampling strategies is provided. We then present the results of our investigation of cohesive conjunction in our corpora, before proceeding to our discussion and conclusions.

### **Cohesion**

Systemic Functional Grammar is employed as the major analytical framework in this chapter because to date it offers the most extensive description of cohesive devices in English and it has been used as a reference for studies both in English and Thai, for example in Baker (2018) and Chanawangsa (1986). Halliday and Matthiessen (2004: 87) identify two resources that constitute the textual component of the linguistic system, namely the structural and the cohesive. They (*ibid.*) describe that how grammar construes structural units up to the rank of the clause complex and stops there; nevertheless, although grammar stops there, semantics does not because the grammar provides cohesion to manage the flow of discourse in order to create semantic links either within or across sentences. Cohesion is therefore the network of lexical, grammatical and other relations that provides links between various parts of the text and these ties organize and to some extent create a text, for example by requiring readers to interpret words and expressions by reference to surrounding sentences or paragraphs (Baker 2018: 194). According to Halliday and Hasan (1976), English cohesive ties can be classified into five types: reference, ellipsis, substitution, conjunction and lexical cohesion. Given its importance to the current research, we home in on conjunction in the following sections.

### **Conjunction**

The term ‘conjunction’ used in the context of cohesion in Systemic Functional Grammar can sometimes cause confusion. According to Bloor and Bloor (2004: 97), conjunction is “the term used to describe the cohesive tie between clauses or sections of text in such a way as to demonstrate a meaningful relationship between them.”

According to Bloor and Bloor, such ties are realized by the use of conjunctive adjuncts. Bloor and Bloor (ibid.) also discuss the difference between ‘conjunctions’ and ‘conjunctive adjuncts’, pointing out that whereas conjunctions are a word class within the same system as noun, verb, adjective, etc., conjunctive adjuncts perform a *function* in the clause and are realized by adverb, nominals and prepositional phrases (ibid.: 57). In other words, although they have a lot in common semantically, they have different grammatical characteristics (ibid.: 98). Conjunctions as a word class are, in turn, split into ‘linkers’ (also known as ‘co-ordinating conjunctions’ or ‘co-ordinators’) and ‘binders’ (also known as ‘subordinating conjunctions’ or ‘subordinators’). The small set of linkers in English includes ‘and’, ‘or’ and ‘but’, while the set of binders is larger and includes such conjunctions as ‘because’, ‘since’, ‘when’, to name just a few. For Bloor and Bloor (2004), conjunction as cohesion is realized by conjunctive adjuncts between sentences, not within sentences.

Halliday and Hasan’s (1976) original work on cohesion identified four types of conjunction in English, namely additive, adversative, causal and temporal. Additive conjunction is described as “a generalized semantic relation in the text-forming component of the semantic system, that is based on the logical notion of ‘and’” (ibid.: 234). It is realized by linkers such as ‘and’, adverbs such as ‘furthermore’, and prepositional phrases such as ‘in addition’, where the latter two function as conjunctive adjuncts. Adversative conjunction encodes the basic meaning of contrary to expectation (ibid.: 250) and is realized by linkers like ‘but’ and binders like ‘though’, and adverbs and prepositional phrases such as ‘however’ and ‘on the other hand’ (ibid.: 242). Causal conjunction indicates relations of cause and effect and is realized by linkers like ‘so’, and adverbs and prepositional phrases such as ‘consequently’ and ‘as a result’ (ibid.: 256). Finally, temporal conjunction indicates, for example, whether clauses refer to events that happen one after the other (i.e., in a sequential relationship) or at the same time (i.e., simultaneously). It is realized by linkers like ‘then’ and adverbs and prepositional phrases such as ‘meanwhile’ or ‘in conclusion’ (ibid.: 266-267).

Based on Halliday and Hasan (1976), Baker (2018: 204) likewise describes how conjunction involves the use of formal markers to relate sentences, clauses and paragraphs to each other and conjunction signals the way the writer or speaker wants the reader or the hearer to relate what is about to be said to what has been said before. Nevertheless, Baker (2018: 205) notices some uncertainty in the literature as to whether

conjunctions that occur within sentences can be considered cohesive. Baker (ibid.) argues that since cohesion is deemed by some linguists to be a relation between sentences rather than within sentences, this means that subordinators are not considered as creating cohesive conjunction. Baker (ibid.) contends, however, that for the purpose of translation, it makes more sense to take a broader view of cohesion by considering any element cohesive as long as it signals a conjunctive-type relation between parts of a text, whether they are sentences, clauses or paragraphs. Baker (ibid.) also points out that the use of conjunction varies across languages, for example, when comparing English and German texts, she (ibid.: 205) finds that English prefers to present information in small chunks and tends to use a wide variety of conjunctions to mark semantic relations between parts of texts while German is likely to express relations through subordination and complex structures.

In addition, the use of conjunction is not only language-specific but it is also sensitive to genre. Smith and Frawley (1983) study the use of conjunctive cohesion in four American English genres: fiction, journalism, religion and science. They find that some genres use more conjunctions than the others. To our knowledge, however, there is no study on the distribution of conjunctions in legal texts compared to other text types. The literature on conjunction in legal texts is mainly concerned with the ambiguity caused by the use of the conjunctions, such as ambiguity caused by the use of ‘and’ and ‘or’ (see Alcaraz and Hughes (2002: 45-46) and Adams and Kaye (2012)).

### **Thai Conjunction**

The literature on Thai conjunction is very limited and rarely cited. The existing literature is mainly composed of linguistic textbooks describing the use of cohesion in the Thai language. Nevertheless, as mentioned in Chapter 1, modern Thai linguistics is heavily influenced by English-language linguistics and we note that the work of Halliday and Hasan (1976) regarding English cohesion is largely employed as the framework to explore Thai cohesion. Thai scholars such as Prasithrathsint (2009) and Sriyapai (2013) thus divide Thai cohesion into four major types (reference, ellipsis, conjunction and lexical cohesion) that are clearly drawn from Systemic Functional Grammar. With respect to the definition, there is also confusion as to whether Thai cohesive conjunction operates within or between sentences. For example, Sriyapai (2013) defines conjunction as a cohesive element that operates between sentences; however, we find that the Thai conjunctions in Sriyapai’s (2013) examples are not only



used within sentences but also used to coordinate, for example, nouns within noun phrases.

Regarding the types of Thai conjunction, Thai scholars have classified Thai conjunction differently, for instance, Chanawangsa (1986) divides Thai conjunction into 16 types whereas Sriyapai (2013) classifies it into 13 types. However, close examination of these types reveals that they can be considered sub-types of Halliday and Hasan's classifications of English conjunction. For example, in Chanawangsa's (1986) classification, a conditional relation is considered a separate type of Thai conjunction, while it is considered as a conjunctive relation under the general heading of causal conjunction within the framework of English-language Systemic Functional Grammar (Halliday and Hasan 1976: 258).

### **Cohesive Conjunction Investigated in this Research**

Within the area of translation studies, there are some studies where the use of conjunction in source texts is compared with that in target texts, and the main focus is on the distribution or frequency of the conjunctions and the factors influencing this distribution (e.g. Biel (2014) and Baker (2018)). Øverås (1998) includes conjunction in her study of explicitation in translation between English and Norwegian (see Chapter 2), and deems insertions into the target texts of conjunctions not present in her source texts to be instances of explicitation, but the data adduced in this source specifically involving conjunction are very sparse, and even then, some relate to coordination within noun phrases rather than between clauses.

There is, to our knowledge, no literature mentioning the use of conjunction in Thai legal texts, whether related to the distribution across genres of conjunctions or the legal translation of conjunction. Therefore, in this section we break new ground by investigating in English to Thai legislative translations and monolingual Thai legislative texts, elements that signal a conjunctive-type relation. We adopt Baker's (2018) standpoint by investigating any conjunctions that link or bind clauses, whether dependent or independent. We investigate only at the clause level for the purpose of consistency. (In other words, we do not consider in our analysis the use of words equivalent to 'and' to coordinate nouns in noun phrases.)

In addition, for the purpose of this thesis, we refer to these instances of conjunction collectively as ‘cohesive conjunction’. As such, cohesive conjunction in this sense can be realized by coordinating and subordinating conjunctions, and adverbial adjuncts, which link, bind or otherwise connect discrete clauses (whether dependent or independent). We investigate the types of conjunction identified by Halliday and Hasan (1976). That said, as already indicated, Halliday and Hasan (1976) recognize four types of conjunction, and see the conditional relation as a sub-type of causal conjunction. In this study we, on the other hand, investigate conditional conjunctions separately since the conditional relation is considered a predominant feature of legal language used to prescribe legal rules (Biel 2014: 144). This chapter therefore examines five relations, namely additive, adversative, temporal, causal and conditional cohesive conjunction.

### **Research Questions**

We set out in this final chapter, to find out:

1. How are instances of cohesive conjunction translated from English into Thai in legislative texts?
2. Are there differences between how cohesive conjunction is typically expressed in legislation translated into Thai and how it is expressed in legislation originally written in Thai?

### **Research Procedure**

The analysis in this chapter contains three parts. In the first part, we investigate the use of cohesive conjunction in the English source texts and their Thai counterparts. We employ the Wordlist function in Sketch Engine (as before, using the modified English Tree Tagger part-of-speech tagset) to create a list of possible cohesive conjunctions in the English source texts. We investigate instances of three tags, namely CC= coordinating conjunction, IN= preposition or subordinating conjunction, and RB=adverb. To find possible conditional cohesive conjunctions, apart from using results derived from employing part-of-speech searches, we conduct a separate search

for conjunctions such as ‘provided that’, which are not picked up by the Wordlist function.<sup>28</sup>

In the second part, we compare the frequency of the Thai translations of the instances of English cohesive conjunction derived above with those in the Thai non-translated corpus. We conduct a search in both corpora and rule out instances that are not clause linkers or binders, or that do not function as conjunctive adjuncts creating cohesion between clauses. We then normalize frequencies of the conjunctions/conjunctive adjuncts across our corpora (by norming to frequency per million words) and calculate log likelihood scores to check if the difference between the results from the two corpora is significant.

In the last part, we compare the frequency of cohesive conjunctions across two sets of texts, namely the English source texts and the Thai monolingual corpus.

As with previous analyses reported on in this thesis, the forms under investigation in this chapter are often polyfunctional: ‘and’, for example, can coordinate nouns in a noun phrase, or it can serve as a coordinating conjunction linking two clauses. As we are interested in only the latter use, we need to manually check the contexts in which ‘and’ occurs, in order to eliminate instances that are irrelevant to our study. As in previous chapters, in cases where there are too many instances of a form to make manual inspection of all of them possible, we generate a random sample of instances (using the Random Sample function in Sketch Engine), and base our judgments on this sample. Our random samples consist of either 10% of the total number of occurrences for the form in question (for extremely frequent forms) or 100 instances of the form, as illustrated respectively by ‘and’ and ‘if’ in Tables 6.2 and 6.6 below. Thus, there are 6,434 instances of ‘and’ in our English source texts (Table 6.2). We manually inspect 643 instances (or 10%) and find only 9 cases where ‘and’ is used as a cohesive conjunction. From this sample we extrapolate that there are 90 instances of ‘and’ used as a cohesive conjunction in the entire English side of our parallel corpus, so we indicate in column 3 of Table 6.2 that the estimated number of uses of ‘and’ as a cohesive conjunction in the corpus is ‘90’ and we norm to a frequency per million

---

<sup>28</sup> This is because ‘provided’ is tagged as a verb (tag: VVD) while ‘that’ is tagged either as a preposition or a subordinating conjunction (tag: IN) or determiner (tag: DT).

(column 4) on this basis. Where such random sampling has been used, this is indicated in column 6 of Table 6.2. Where the cell in column 6 has been left empty, this indicates that *all* instances of the form in question were manually inspected.

Likewise, in our comparisons between translated Thai and non-translated Thai, random sampling is used to make manual inspection of data practicable. Table 6.7 below, for example, presents results regarding the Thai form *และ* /lae/ (and), among others. There are actually 22,887 instances of this form in our Thai monolingual corpus, and 6,308 instances in our Thai translation corpus. We inspect a random sample of 10% of instances (that is, 2,288 and 630 instances respectively) and identify 199 and 9 instances respectively where *และ* /lae/ is used as a cohesive conjunction linking clauses. On this basis we extrapolate to a total of 1,990 and 90 uses of *และ* /lae/ as a cohesive conjunction linking clauses in the Thai monolingual and Thai translation corpora respectively.

## Results and Analysis

In the following, our results are presented for each of the three stages described above.

### *Cohesive Conjunction in the English-Thai Parallel Corpus*

After we find the list of possible cohesive conjunctions/conjunctive adjuncts, we inspect all instances in each tag category, and rule out instances that are obviously not instances of cohesive conjunction. There are many instances where the form in question is used as a non-cohesive preposition; this is the case, for example, with all instances of ‘whether’. We then group candidate instances of cohesive conjunction into the five categories previously mentioned, the details of which are shown in Table 6.1.

Type of conjunction	Candidate Instances of Cohesive Conjunction (total number of occurrences)
Additive	and (6,434), also (145), additionally(1), either (33), furthermore (4), moreover (1), neither (3), nor (21), or (3,123)
Adversative	but (56), even though (1), although (2), notwithstanding (26), nonetheless (2), nevertheless (8), however (34)
Causal	because (7), accordingly (1), therefore (7), thus (11), so (104)
Temporal	after (289), as (1,368), before (96), while (16), then (16)
Conditional	if (352), if..then (3), provided that (43), unless (97)

**Table 6.1: Candidate instances of cohesive conjunction**

We then filter these instances thoroughly and investigate only the ones that create cohesion between clauses, and examine how these instances of cohesive conjunction in the five categories are translated into Thai. The results of this investigation are as follows.

#### Additive Cohesive Conjunction

After filtering candidates in the additive category, we find that there are instances of cohesive conjunction realized by five forms, namely ‘and’, ‘or’, ‘additionally’, ‘furthermore’, and ‘moreover’, while there are no cohesive instances of ‘also’, ‘either’, ‘neither’ or ‘nor’. The cohesive uses of these forms and their translations into Thai are addressed below.

## And

There are 6,434 instances of ‘and’ which are tagged as coordinating conjunctions. A random sample of 10% of all instances, i.e., 643 instances, was created. We find that only 9 instances of ‘and’ function as clause linkers. When we examine the translations of these instances, all instances of ‘and’ are translated as และ /lae/ (and), which is considered the closest equivalent of ‘and’ in Thai. Such a translation is illustrated in Example 6.1.

### *Example (6.1)*

English source text:

In determining such adjustment, due regard shall be had to the other provisions of this Convention **and** the competent authorities of the Contracting States shall if necessary consult each other.

Thai translation:

ในการกำหนดการปรับปรุงเช่นว่านั้นต้องคำนึงถึงบทบัญญัติอื่น ๆ แห่งอนุสัญญานี้ด้วยตามควร**และ**หากจำเป็นให้เจ้าหน้าที่ผู้มีอำนาจของ รัฐผู้ทำสัญญาทั้งสองปรึกษาหารือซึ่งกันและกัน

Back translation (gloss):

in determination improvement such must consider other provision of convention this **/lae/** (and) if necessary hai authority competent of state contracting both consult each other.

## Or

There are 3,123 instances of ‘or’ tagged as a coordinating conjunction. Our random sample thus contains 10% or 312 instances. We find that only one instance of ‘or’ is used to link between clauses, and it is translated, as expected, by หรือ /rue/ (or). See Example 6.2.

### *Example (6.2)*

English source text:

Each Party shall not allow, by taking appropriate measures, the manufacture, import or export of mercury-added products listed in Part I of Annex A after the phase-out date specified for those products, ...*or* the Party has a registered exemption pursuant to Article 6.

Thai translation:

ไม่ให้แต่ละภาคีอนุญาตให้มีการผลิต นำเข้าหรือส่งออกผลิตภัณฑ์ที่เติมปรอท ตามที่มี รายชื่อในส่วนที่ ๑ ของภาคผนวก  
เอ ภายหลังจากที่กำหนดให้มีการเลิก ตามที่ระบุไว้ ...*หรือ*กรณีที่ภาคีนั้นได้จดทะเบียนขอยกเว้นตามข้อ ๖

Back translation (gloss):

Not hai each party allow manufacture import or export product add mercury list  
in part 1 of annex A after date specify to phase out */rue/* (or) in case party that  
register exemption pursuant to article 6

Additionally, Moreover and Furthermore

We find 1 instance of ‘additionally’ and verify that it is used as a conjunctive adjunct,  
where its translation is นอกจากนี้<sup>๓</sup> /nok/+/chak/+/ni/ (additionally/moreover/furthermore) as  
shown in Example 6.3.

Also, we find one instance of ‘moreover’ and four instances of ‘furthermore’ which also  
function as conjunctive adjuncts. The translators translate these words as นอกจากนี้<sup>๓</sup>  
/nok/+/chak/+/ni/ (additionally/moreover/furthermore),

นอกเหนือจากนี้<sup>๓</sup> /nok/+/nuea/+/chak/+/ni/ (additionally/moreover/furthermore),

ยิ่งไปกว่านั้น<sup>๓</sup> /ying/+/pai/+/kwa/+/nan/ (additionally/moreover/furthermore) and

ยิ่งกว่านั้น<sup>๓</sup> /ying/+/kwa/+/nan/ (additionally/moreover/furthermore).

Although these four Thai words look different, they have very similar meaning, and are  
used when a speaker needs to add information in Thai. As such, they all can be  
considered close equivalents of their English counterparts.

### **Example (6.3)**

English source text:

The information submitted under Article 7, paragraph 1, by each Party included in Annex I shall be reviewed as part of the annual compilation and accounting of emissions inventories and assigned amounts. *Additionally*, the information submitted under Article 7, paragraph 2, by each Party included in Annex I shall be reviewed as part of the review of communications.

Thai translation:

ให้ข้อมูลข่าวสารที่นำเสนอภายใต้มาตรา 7วรรค 1 โดยประเทศภาคีในภาคผนวกที่ I ได้รับการทบทวนในฐานะที่เป็นส่วนหนึ่งของการรวบรวมและคิดบัญชีปริมาณการปล่อยก๊าซและปริมาณที่กำหนดรายปี นอกจากนี้ให้ข้อมูลข่าวสารที่นำเสนอภายใต้มาตรา 7 วรรค 2 โดยประเทศภาคีแต่ละประเทศภาคีในภาคผนวกที่ I ได้รับการทบทวนในฐานะที่เป็นส่วนหนึ่งของการทบทวนรายงานแห่งชาติ

Back translation (gloss):

Hai information submitted under article 7 paragraph 1 by state party in annex I receive review as part of compilation and accounting of emission gas and amount assigned annually */nok/+/chak/+/ni/* (additionally/moreover/furthermore) hai information submitted under article 7 paragraph 2 by state party each in annex I receive review as part of review national

Based on the findings above, we notice the low frequency of additive cohesive conjunction in the English source texts. In addition, the translations of these cohesive conjunctions/conjunctive adjuncts tend to be predictable. Table 6.2 presents the details.



English additive conjunctions	Total number of occurrences	(Estimated) Number of cohesive conjunctive uses	Frequency per million (corpus size: 172,739 words)	Thai translation	Note
and	6,434	90	521	และ /lae/	10% random sample
or	3,123	10	57.89	หรือ /rue/	10% random sample
additionally	1	1	5.78	นอกจากนี้ /nok/+/chak/ +/ni/	-
furthermore	4	4	23.15	นอกเหนือจากนี้ /nok/+/nuea/ +/chak/+/ni/ , ยิ่งไปกว่านั้น /ying/+/pai/ +/kwa/+/nan /ยิ่งกว่านั้น /ying/+/kwa +/nan/	-
moreover	1	1	5.78	นอกจากนี้ /nok/+/chak/ +/ni/	-
total		106			

**Table 6.2: Additive cohesive conjunction**

#### Adversative Cohesive Conjunction

The number of instances considered cohesive conjunctions/conjunctive adjuncts is low in the adversative category. For example, all 26 instances of ‘notwithstanding’ are prepositions that do not bind clauses, so they are excluded from our investigation. There are five cohesive conjunctions under consideration.

#### But

There are 56 instances of ‘but’, of which 17 instances function as clause linkers. Fifteen out of 17 instances of ‘but’ are translated as แต่ /tae/, the closest equivalent of ‘but’ in Thai (Example 6.4), while the remaining two instances are omitted in Thai translation (Example 6.5).

#### *Example (6.4)*

English source text:

A declaration may be withdrawn at any time by notification to the Secretary-General, **but** such a withdrawal shall not affect communications pending before the Committee.

Thai translation:

รัฐภาคีที่เกี่ยวข้องอาจเพิกถอนการประกาศนี้ได้ทุกเมื่อ โดยการแจ้งต่อเลขาธิการสหประชาชาติแต่การเพิกถอนดังกล่าวจะต้องไม่กระทบต่อการร้องเรียน ซึ่งอยู่ในระหว่างการพิจารณาของคณะกรรมการ

Back translation (gloss):

state party related might withdraw notification this any time by notification to secretary-general /*tae*/ (but) withdrawal such shall not affect communication pending consideration of committee

#### **Example (6.5)**

English source text:

The Committee shall confidentially bring any communication referred to it to the attention of the State Party alleged to be violating any provision of this Convention, **but** the identity of the individual or groups of individuals concerned shall not be revealed without his or their express consent.

Thai translation:

คณะกรรมการจะแจ้งข้อร้องเรียนที่ได้รับต่อรัฐภาคีที่ถูกกล่าวหาว่าละเมิดข้อบทของอนุสัญญา โดยจะไม่เปิดเผยชื่อบุคคลหรือกลุ่มบุคคลที่เกี่ยวข้องโดยมิได้รับอนุญาต

Back translation (gloss):

committee will bring communication received to state party which thuk allege violate provision of convention (**omission**) by not revealing name of person or group person which related without permission

Although, Even though, Nevertheless, Nonetheless

In addition, we find that one out of two instances of ‘although’, one instance of ‘even though’ and all instances of ‘nevertheless’ (8) and ‘nonetheless’ (2) are cohesive conjunctions/conjunctive adjuncts, and their Thai translations conform to the English

source text where all instances are translated into Thai conjunctions/conjunctive adjuncts, namely อย่างไรก็ตาม /yang/+/rai/+/ko/+/tam/, แม้ว่า /mae/+/wa/, อย่างไรก็ตาม /yang/+/rai/+/ko/+/di/, แม้กระนั้น /mae/+/kra/+/nan/; all are close equivalents of their English counterparts (as can be seen in Example 6.6).

**Example (6.6)**

English source text:

Any State Party to the Convention which is in arrears with the payment of its compulsory or voluntary contribution for the current year and the calendar year immediately preceding it shall not be eligible as a Member of the World Heritage Committee, *although* this provision shall not apply to the first election.

Thai translation:

รัฐภาคีแห่งอนุสัญญาฯ ใดค้างชำระค่าบำรุงหรือเงินบริจาคประจำปีภายในปีปฏิทินนั้น จะไม่มีสิทธิ์สมัครเข้ารับการเลือกตั้งเป็นกรรมการในคณะกรรมการมรดกโลก อย่างไรก็ตาม ความค้างชำระข้างต้นไม่มีผลบังคับใช้กับการเลือกตั้งในครั้งแรก

Back translation (gloss):

state party of convention any in arrear with payment contribution or donation annually within year calendar that will not eligible to apply for member of world heritage committee, /yang/+/rai/+/ko/+/tam/ (although) provision this does not apply to election first

However

Thirty three out of 34 instances of ‘however’ are clause binders. There are two cases where ‘however’ is omitted in Thai translation while the remaining instances are translated by their closest equivalents, namely อย่างไรก็ตาม /yang/+/rai/+/ko/+/tam/, แม้ว่า /mae/+/wa/, and อย่างไรก็ตาม /yang/+/rai/+/ko/+/di/. The example where the translators omit the translation of ‘however’ in Thai is as shown in Example 6.7 below.

**Example (6.7)**

English source text:

In no case shall the compulsory contribution of States Parties to the Convention exceed 1% of the contribution to the regular budget of the United Nations Educational, Scientific and Cultural Organization. *However*, each State...may declare,..., acceptance or accession,...

Thai translation:

หากไม่มีกรณีเป็นอย่างอื่น ค่าบำรุงประจำปีของรัฐบาลจะเป็นจำนวนเท่ากับร้อยละ ๑ ของเงินอุดหนุนประจำปีที่รัฐบาลมอบให้กับองค์การการศึกษา วิทยาศาสตร์และวัฒนธรรมแห่งสหประชาชาติ โดยหลักปฏิบัติทั่วไป รัฐบาลทุกรัฐที่... จะต้องประกาศ,..., การให้สัตยาบัน หรือการรับรอง

Back translation (gloss):

If no case other contribution annual of state party be equal to 1%<sup>29</sup> of contribution annual which state party give to organization science and culture of united nation based on general practice (*omission*) state party every state...may declare,..., acceptance or accession...

Again we notice a low frequency of adversative cohesive conjunction used in the English legal texts, and the majority of their Thai translations adhere to their English counterparts, with a few exceptions where counterparts for the conjunctions ‘but’ and ‘however’ are omitted in the Thai translations. These results are summarized in Table 6.3.

---

<sup>29</sup> The translation in this case appears to be based on a misinterpretation.

English adversative conjunctions	Total number of occurrences	Number of cohesive conjunctive uses	Frequency per million (corpus size: 172,739 words)	Thai translation
but	56	17	98.41	แต่ /tae/, omission
even though	1	1	5.78	อย่างไรก็ตาม /yang/+/rai/+/ko/+/tam/
although	2	1	5.78	อย่างไรก็ดี /yang/+/rai/+/ko/+/di/
nonetheless	2	2	11.57	อย่างไรก็ตาม /yang/+/rai/+/ko/+/tam/, อย่างไรก็ดี /yang/+/rai/+/ko/+/di/
nevertheless	8	8	46.31	อย่างไรก็ตาม /yang/+/rai/+/ko/+/tam/, แม้ว่า /mae/+/wa/
however	34	33	191.03	อย่างไรก็ตาม /yang/+/rai/+/ko/+/tam/, แม้ว่า /mae/+/wa/, อย่างไรก็ดี /yang/+/rai/+/ko/+/di/, แม้กระนั้น /mae/+/kra/+/nan/
total		62		

**Table 6.3: Adversative cohesive conjunction**

#### Temporal Cohesive Conjunction

We filter the candidates for temporal cohesive conjunction, and find instances of four conjunctions which link or bind, namely ‘after’, ‘before’, ‘while’ and ‘then’.

#### After

There are 289 instances of ‘after’. When we investigate all lines, we find that 35 instances of ‘after’ are used as clause-binding, subordinating conjunctions, while the rest function as prepositions. The majority of these 35 instances of ‘after’ are translated by the closest equivalents of ‘after’, namely หลังจาก /lung/+/chak/ (20 instances), ภายหลัง /pai/+/lung/ (5 instances) and ถัดจาก /that/+/chak/ (2 instances), as illustrated in Example 6.8 below.

### **Example (6.8)**

English source text:

*After* such proceedings have been completed with regard to an inquiry made in accordance with paragraph 2, the Committee may, after consultations with the State Party concerned, decide to include a summary account of the results of the proceedings...

Thai translation:

หลังจากที่การดำเนินกระบวนการพิจารณาเห็นว่าสิ้นสุดลงในส่วนที่เกี่ยวกับการไต่สวนที่กระทำตามวรรค ๒ แล้ว คณะกรรมการฯ สามารถวินิจฉัยหลังจากการหารือกับรัฐภาคีที่เกี่ยวข้อง ให้รวมเนื้อหาโดยสรุปของผลการดำเนินกระบวนการพิจารณา...

Back translation (gloss):

*/lung/+chak/* (after) proceeding such complete with regard to inquiry make in accordance with paragraph 2 committee may decide after consultation with state party concerned to include account summary of results of proceedings...

Nevertheless, there are some instances where the translators prefer other words that are considered somewhat equivalent to ‘after’. We find that two instances are translated by *ก็ต่อเมื่อ* */ko/+to/+muea/*, an equivalent of ‘only if’. When the translators translate ‘after’ as *ก็ต่อเมื่อ* */ko/+to/+muea/* (only if), we notice a shift from a temporal relation to a conditional relation, an example of which is given in Example 6.9 below.

### **Example (6.9)**

English source text:

The Committee shall deal with a matter referred to it in accordance with paragraph 2 of this article *after* it has ascertained that all available domestic remedies have been invoked and exhausted in the case,...

Thai translation:

คณะกรรมการจะจัดการกับกรณีอ้างอิงตามวรรค 2 ของข้อนี้ *ก็ต่อเมื่อ* ได้พิจารณาอย่างชัดเจนแล้วว่าได้มีการใช้มาตรการภายในประเทศทั้งหมดแล้ว แต่ไม่ประสบผลในกรณีดังกล่าว...

Back translation (gloss):

Committee will deal with matter refer to in accordance with paragraph 2 of article this */ko/+to/+muea/* (only if) consider clearly that use remedy domestic all but not effective

In addition, there are two instances of ‘after’ being translated as นับจาก */nup/+chak/* and นับแต่ */nup/+tae/*, where their meanings are equivalent to that of ‘since’ (Example 6.10).

### **Example (6.10)**

English source text:

...the denunciation takes effect upon the expiration of such longer period *after* the notification is received by the depository.

Thai translation:

การบอกเลิกมีผลเมื่อสิ้นสุดระยะเวลาที่นานกว่านั้นนับแต่ผู้ดูแลอนุสัญญาได้รับคำบอกกล่าว

Back translation (gloss):

Denunciation have effect when the end of period longer such */nup/+tae/* (since) depository receive notification

In English, there can be confusion as to whether ‘after’ and ‘since’ are interchangeable. According to the Cambridge Dictionary, as a conjunction, ‘after’ means “at a time that is later than another event” while ‘since’ means “from a particular time in the past until a later time, or until now”. These meanings can also be applied to the meaning of Thai words หลังจาก */lung/+chak/* (after) and นับแต่ */nup/+tae/* (since). Based on these definitions, we can say that there is a slight difference in the starting point of Example 6.10 because ‘after’ starts at a particular specified point, but ‘since’ includes the entire interval from then until now.

## Before

There are 96 instances of ‘before’, but only 2 instances are clause binders and these two instances are translated as ก่อน /kon/, the closest equivalent of ‘before’ as presented in Example 6.11.

### *Example (6.11)*

English source text:

The consignor must furnish such information and such documents as are necessary to meet the formalities of customs, police and any other public authorities **before** the cargo can be delivered to the consignee.

Thai translation:

ผู้ตราส่งต้องให้ข้อมูลและเอกสารที่จำเป็นต่อการปฏิบัติตามพิธีการทางศุลกากร ทางการค้า และพิธีการของหน่วยงานของรัฐหน่วยงานอื่น **ก่อน**ที่จะส่งมอบของให้แก่ผู้รับตราส่งได้

Back translation (gloss):

Consignor must furnish information and document as necessary to formalities customs police and authority public other **/kon/** (before) deliver cargo to consignee

## While

Five out of 16 instances of ‘while’ function as temporal clause binders and again they are all translated by the closest equivalent, ในขณะที่ /nai+/kha+/na+/ti/, as presented in Example 6.12 below.

### *Example (6.12)*

English source text:

..,the States Parties concerned shall be entitled to send a representative to take part in the proceedings of the Committee, without voting rights, **while** the matter is under consideration.

Thai translation:



รัฐภาคีที่เกี่ยวข้อง มีสิทธิอันชอบธรรมที่จะจัดส่งผู้แทนเข้าร่วมในกระบวนการพิจารณาของคณะกรรมการ โดยไม่มีสิทธิลงคะแนนเสียง*ในขณะที่*คณะกรรมการพิจารณาเรื่องดังกล่าว

Back translation (gloss):

State party concerned entitle to send representative to take part in proceeding of committee without right to vote */nai/+/kha/+/na/+/ti/* (while) committee decide about matter this.

Then

Three out of 17 cohesive instances of ‘then’ connect clauses in our English source texts, two of which are translated, as expected, by จากนั้น */chak/+/nan/* (then/after) (Example 6.13), while one instance is omitted in the Thai translation (Example 6.14).

#### *Example (6.13)*

English source text:

The Secretary-General of the United Nations shall prepare a list in alphabetical order of the persons thus nominated and shall submit it to the States Parties to the present Covenant. The election to fill the vacancy shall **then** take place in accordance with the relevant provisions of this part of the present Covenant.

Thai translation:

เลขาธิการสหประชาชาติต้องจัดทำรายนามบุคคลทุกคนซึ่งได้รับการเสนอนามตามลำดับอักษร และต้องเสนอรายนามดังกล่าวไปยังรัฐภาคีแห่งกติกานี้ จากนั้นให้มีการเลือกตั้งเพื่อแทนตำแหน่งที่ว่างตามบทบัญญัติที่เกี่ยวข้องในภาคนี้ของกติกานี้

Back translation (gloss):

secretary-general of united nation shall prepare name list all person nominated in order alphabet and shall submit name list such to state party to convention this */chak/+/nan/* (then) hai have election to replace position vacant in accordance with provision this of covenant this

### Example (6.14)

English source text:

If a State Party considers that another state Party is not giving effect to the provisions of this Convention, it may bring the matter to the attention of the Committee. The Committee shall *then* transmit the communication to the state Party concerned.

Thai translation:

หากรัฐภาคีใดเห็นว่ารัฐภาคีอื่นไม่ปฏิบัติตามข้อบทของอนุสัญญานี้ รัฐภาคีนั้นอาจนำเรื่องเข้าสู่การพิจารณาของ  
คณะกรรมการ คณะกรรมการจะส่งต่อเรื่องดังกล่าวไปยังรัฐภาคีที่เกี่ยวข้องนั้น

Back translation (gloss):

if state party consider that state party another do not follow the provision of convention this state party such may bring matter to consideration of committee (*omission*) committee will transmit communication such to state party concerned

A summary of the translation of temporal cohesive conjunctions/conjunctive adjuncts is presented in Table 6.4.

Temporal conjunctions	Total number of occurrences	Number of cohesive conjunctive uses	Frequency per million (corpus size: 172,739 words)	Thai translation
after	289	35	202.61	หลังจาก /lung/+/chak/, ภายหลัง /pai/+/lung/, ถัดจาก /that/+/chak/, ก็ต่อเมื่อ /ko/+/to/+/muea/, นับจาก /nup/+/chak/ and นับแต่ /nup/+/tae/
before	96	2	11.57	ก่อน /kon/
while	16	5	28.94	ในขณะที่ /nai/+/kha/+/na/+/ti/
then	17	3	17.36	จากนั้น /chak/+/nan/, omission
total		45		

Table 6.4: Temporal cohesive conjunction

## Causal Cohesive Conjunction

After we filter all candidates, there are three causal cohesive conjunctions, ‘because’, ‘accordingly’ and ‘therefore’.

### Because

Two out of seven instances of ‘because’ are used to bind clauses, and they are both translated by their closest equivalent in Thai, เพราะว่า /pro/+wa/ (because) (Example 6.15).

#### *Example (6.15)*

English source text:

...it does not lose its international character merely *because* one contract or a series of contracts is to be performed entirely within the territory of the same State.

Thai translation:

...ยอมไม่สูญเสียลักษณะระหว่างประเทศ เพียง*เพราะว่า*จะต้องปฏิบัติการทั้งหมดภายในอาณาเขตของรัฐเดียวกันตามสัญญาฉบับเดียวหรือสัญญาเป็นชุด

Back translation (gloss):

...not lose character international merely /*pro*+/wa/ (because) shall perform entirely within territory of state same in accordance with contract one or contract serie

### Accordingly and Therefore

Only two out of 10 instances of ‘accordingly’ indicate cohesive conjunction and they are translated by ด้วยเหตุนี้ /duai+/hed+/ni/ and ในกรณีนี้ /nai+/kan+/ni/, both of which are standard translations of ‘accordingly’ in Thai (Example 6.16). Regarding ‘therefore’, only two instances are cohesive conjunctions, and both of them are in a preamble. Also, these two instances are translated by the Thai closest equivalent of ‘therefore’, ฉะนั้น /cha+/nan/ (Example 6.17).

#### *Example (6.16)*

English source text:

*Accordingly*, a request for extradition based on such an offence may not be refused on these grounds alone.

Thai translation:

ด้วยเหตุนี้ ถ้าขอส่งผู้ร้ายข้ามแดนซึ่งตั้งอยู่บนความผิดดังกล่าวไม่สามารถได้รับการปฏิเสธเพราะเหตุผลดังกล่าวเพียงเท่านั้น

Back translation (gloss):

/duai/+/hed/+/ni/ (accordingly), request for extradition based on offence such cannot receive refusal because of ground such only

#### **Example (6.17)**

English source text:

Now, *therefore*, THE GENERAL ASSEMBLY proclaims THIS UNIVERSAL DECLARATION OF HUMAN RIGHTS as a common standard of achievement for all peoples and all nations,

Thai translation:

ฉะนั้น บัดนี้ สมัชชาจึงประกาศปณิญาสากลว่าด้วยสิทธิมนุษยชนนี้ ให้เป็นมาตรฐานร่วมกันแห่งความสำเร็จสำหรับประชาชนทั้งมวล และประชาชาติทั้งหลาย

Back translation (gloss):

/cha/+/nan/ (therefore), now general assembly proclaim declaration universal of right human as standard common of achievement for people all and nation all

A summary of the translation of causal cohesive conjunctions/conjunctive adjuncts is presented in Table 6.5.

Causal conjunctions	Total number of occurrences	Number of cohesive conjunctive used	Frequency per million (corpus size: 172,739 words)	Thai translation
because	7	2	11.57	เพราะว่า /pro/+/wa/
accordingly	10	2	11.57	ด้วยเหตุนี้ /duai/+/hed/+/ni/, ในการนี้ /nai/+/kan/+/ni/
therefore	6	2	11.57	ฉะนั้น /cha/+/nan/
total		6		

**Table 6.5: Causal cohesive conjunction**

#### Conditional Cohesive Conjunction

Conjunctions expressing conditional relations fall under the category of English causal conjunction within the framework of Systemic Functional Grammar (Halliday and Hasan 1976). Nevertheless, Chanawangsa (1986) considers conjunctions used in conditional relations as a separate type of Thai conjunction. Conditional relations are also considered a distinctive feature of legislative texts and play an important role in formulating legal rules, especially those following the ‘if-then’ pattern (Biel 2014: 144). We thus expect to see a high frequency of this type of conjunction in our English source texts and so decide to investigate the English-Thai translation of conditional conjunctions separately. After thorough examination of all instances, we find four cohesive conditional conjunctions, namely ‘if’, ‘if-then’, ‘provided that’, and ‘unless’.

#### If

We find 349 instances of ‘if’. We take a random sample of 100 instances, 84 of which turn out to be cohesive conjunctions. Therefore, the estimated number of instances of ‘if’ used as a cohesive conjunction is 293. We then investigate how ‘if’ is translated into Thai, and we find that the translators translated ‘if’ as ถ้า /tha/, หาก /hak/, or ถ้าหาก /tha/+/hak/ (all fairly standard Thai equivalents of ‘if’). See Example 6.18 below.

#### **Example (6.18)**

English source text:

*If* a State makes no declaration under this Article, the Convention is to extend to all territorial units of that State.

Thai translation:

หากรัฐไม่ทำคำประกาศตามข้อนี้ อนุสัญญาย่อมมีผลครอบคลุมทุกหน่วยดินแดนของรัฐนั้น

Back translation (gloss):

*/hak/* (if) state make no declaration under article this convention must extend to all unit territory of state that

If-then

We find three instances of the ‘if-then’ pattern. Two of them are translated as ถ้า */tha/*, which is considered the closest equivalent of ‘if-then’. Nevertheless, we find one example where the translator uses ถ้าหาก */tha/+hak/* (an equivalent of ‘if’), but adds the word จากนั้น */chak/+nan/* which literally means ‘after that’, and which is a close equivalent of ‘then’.

### *Example (6.19)*

English source text:

*If* the Conference of the Parties decides that the proposal shall proceed, the Committee shall *then* prepare the risk management evaluation.

Thai translation:

ถ้าหากที่ประชุมรัฐภาคีมีมติว่า ข้อเสนอจะต้องดำเนินต่อไป จากนั้นให้คณะกรรมการฯ เตรียมการ สำหรับการตรวจสอบการจัดการความเสี่ยง

Back translation (gloss):

*/tha/+hak/* (if) conference state party decide that proposal shall proceed  
*/chak/+nan/* (then) hai committee prepare for evaluation management risk

Provided that

There are 43 instances of ‘provided that’ followed by a clause, more than half of which (27 instances) are translated by their close equivalents, which are หาก */hak/* (if), ถ้า */tha/* (if) โดยมีเงื่อนไขว่า */doi/+mi/+nguean/+khai/+wa/* (under the condition that), and ภายใต้เงื่อนไขว่า */phai/+tai/+nguean/+khai/+wa/* (under the condition that). In addition, we find five instances of ‘provided that’ translated by words which are not considered close

equivalents, but still express a conditional relation, namely เมื่อ/muea/ (when), เว้นแต่ /wen/+/tae/(unless), and ตราบเท่าที่ /trap/+/tao/+/ti/(as long as). Example 6.20 illustrates the translation of ‘provided that’ as Thai เว้นแต่ /wen/+/tae/ (unless).

**Example (6.20)**

English source text:

Nothing in this Convention shall prevent the parties in the case of combined carriage from inserting in the document of air carriage conditions relating to other modes of carriage, **provided that** the provisions of this Convention are observed as regards the carriage by air.

Thai translation:

ไม่มีบทบัญญัติใดในอนุสัญญานี้ซึ่งขัดแย้งกับสัญญาในกรณีของการรับขนร่วมกัน ในการระบุนเงื่อนไขเกี่ยวกับการรับขน โดยแบบอื่นไว้ในเอกสารการรับขนทางอากาศ **เว้นแต่** ได้มีการปฏิบัติตามบทบัญญัติแห่งอนุสัญญานี้ในส่วนที่เกี่ยวกับการรับขนทางอากาศ

Back translation (gloss):

no provision any in convention this prevent party in case of carriage combined in inserting condition relating to carriage by mode other in document carriage air /wen/+/tae/ (unless) have observation in accordance with provision of convention this relating to carriage air

We find four instances where the translators omit the translation of ‘provided that’ in Thai version (Example 6.21).

**Example (6.21)**

English source text:

The Committee may accept contributions to be used only for a certain programme or project, **provided that** the Committee shall have decided on the implementation of such programme or project.

Thai translation:

คณะกรรมการมรดกโลกจะรับและใช้เงินกองทุนในแผนงานหรือโครงการที่ได้จัดเตรียมรายละเอียดการดำเนินงานที่ชัดเจน เพียงพอสำหรับการพิจารณาตัดสินใจเท่านั้น

Back translation (gloss):

committee heritage world will receive and use fund in programme or project (*omission*) which prepare detail of implementation clearly enough for decision making only

There are six instances where <sup>ทั้งนี้</sup> /tang/+/ni/ is used in Thai translated version. According to the Thai Royal Institute Dictionary B.E.2554, <sup>ทั้งนี้</sup> /tang/+/ni/ is a Thai conjunction used to refer back to the previous statement. It is difficult to find an equivalent of <sup>ทั้งนี้</sup> in English, but in the current context, it could be back-translated as ‘so as to’. Considering that the conditional relation is a sub-type of causal relation, when the English ‘provided that’ is translated by <sup>ทั้งนี้</sup> /tang/+/ni/ (so as to), it can be considered that the Thai translation still conforms somewhat to its source text, as illustrated in Example 6.22.

**Example (6.22)**

English source text:

Amendments to Annexes A and B to this Protocol shall be adopted and enter into force in accordance with the procedure set out in Article 20, *provided that* any amendment to Annex B shall be adopted only with the written consent of the Party concerned.

Thai translation:

การแก้ไขภาคผนวก ก และ ข ของพิธีสารนี้จะต้องมีการรับรองและมีผลบังคับใช้ตามขั้นตอนที่ได้กำหนดไว้ในมาตรา 20 <sup>ทั้งนี้</sup> การปรับปรุงแก้ไขใดๆในภาคผนวก ก และ ข ต้องได้รับการรับรองเป็นลายลักษณ์อักษรจากประเทศภาคีที่เกี่ยวข้อง เท่านั้น

Back translation (gloss):



amendment to annex a and b of protocol this shall have adoption and enforce in accordance with procedure set out in Article 20 */tang/+/ni/* amendment any to annex a and b shall receive adoption in writing from state party related only

Moreover, there is one instance where แต่ */tae/* (but) is used, which indicates a shift from a conditional to an adversative relation.

**Example (6.23)**

English source text:

Extraordinary meetings of the Conference of the Parties shall be held at such other times as may be deemed necessary by the Conference, or at the written request of any Party *provided that* it is supported by at least one third of the Parties.

Thai translation:

การประชุมพิเศษของที่ประชุมใหญ่ภาคีจะต้องจัดขึ้นในช่วงเวลาอื่นตามที่ที่ประชุมใหญ่เห็นชอบ หรือในเวลาที่เหมาะสม  
ร้องขอเป็นลายลักษณ์อักษร แต่ต้องได้รับการสนับสนุนจากภาคีทั้งหมดอย่างน้อยหนึ่งในสาม

Back translation (gloss):

meeting extraordinary of conference of party shall hold in time other which meeting approve or in time which party any request written */tae/* (but) shall receive support from state at least one third.

Like ‘if’, ‘provided that’ signals a conditional relation where the subordinating clause describes the condition and the main clause has normative content. In cases where there is a shift from a conditional relation to another relation, this might cause the normative content to change.

Unless

There are 97 instances of ‘unless’, all of which are clause binders. Almost all instances (96) are translated by fairly standard equivalents, namely เว้นแต่ */wen/+/tae/* (unless/except), เว้นไว้แต่ */wen/+/wai/+/tae/* (unless/except), ยกเว้นแต่ */yok/+/wen/+/tae/* (unless/except), เว้นเสียแต่ */wen/+/sia/+/tae/* (unless/except), ยกเว้น */yok/+/wen/* (except),

นอกจาก/nok/+/chak/ (except), นอกเสียจาก/nok/+/sia/+/chak/ (except) and หาก...ไม่ /hak/.../mai/ (if...not). There is one instance (Example 6.24) where ‘unless’ is translated into ตราบใดที่...ไม่ /trap/+/dai/+/ti/.../mai/ (as long as...not); but it is considered similar to other equivalents mentioned earlier.

**Example (6.24)**

English source text:

The expiry of a declaration, a notice of revocation or a new declaration shall not in any way affect proceedings pending before an arbitral tribunal or the International Court of Justice *unless* the parties to the dispute otherwise agree.

Thai translation:

การสิ้นสุดการมีผลของคำประกาศ การแจ้งระงับการมีผลหรือการแจ้งคำประกาศใหม่จักต้องไม่มีผลต่อกระบวนการที่กำลัง  
รอการพิจารณาจากอนุญาโตตุลาการ หรือศาลยุติธรรมระหว่างประเทศ *ตราบใดที่*คู่กรณี*ไม่*ตกลงเป็นอย่างอื่น

Back translation (gloss):

expiration effect of declaration notice of revocation or declaration new shall not affect proceeding pending before arbitration or court justice /trap/+/dai/+/ti/ (as long as) party to the dispute /mai/ (not) agree otherwise

The total number of occurrences of these conditional cohesive conjunctions and their Thai translations is presented in Table 6.6.

Conditional conjunctions	Total number of occurrences	(Estimated) Number of cohesive conjunctive uses	Frequency per million (corpus size: 172,739 words)	Thai translation	Note
If	349	293	1,696.20	ถ้า /tha/, หาก /hak/, ถ้าหาก /tha/+/hak/	random sample of 100
If-then	3	3	17.36	ถ้า /tha/, ถ้าหาก /tha/+/hak/, ถ้าหาก...จากนั้น /tha/+/hak/.../cha k/+/nan/	-
Provided that	43	43	248.93	หาก /hak/, ถ้า /tha/, โดยมีเงื่อนไขว่า /doi/+/mi/+/ngue an/+/khai/+/wa/ , ภายใต้เงื่อนไขว่า /phai/+/tai/+/ngu ean/+/khai/+/wa/, เมื่อ/muea/, เว้นแต่ /wen/+/tae/, ตราบเท่าที่ /trap/+/tao/+/ti/, ทั้งนี้ /tang/+/ni/, แต่/tae/ ,omission	-
Unless	97	97	561.54	เว้นแต่/wen/+/tae/, เว้นไว้แต่ /wen/+/wai/+/tae/ , ยกเว้น แต่/yok/+/wen/+/t ae/, เว้นเสียแต่/wen/+/sia /+/tae/, ยกเว้น /yok/+/wen/, นอกจาก /nok/+/chak/, นอกเสีย จาก/nok/+/sia/+/ch ak/,หาก...ไม่ /haak...mai/, ตราบ ใดที่...ไม่ /trap/+/dai/+/ti/... mai/	-
Total		436			

**Table 6.6: Conditional cohesive conjunction**

### Summary of Findings on Cohesive Conjunction in Source and Target Texts

The investigation so far has shown that conditionals constitute by far the most frequent type of cohesive conjunction in our source texts (with an estimated 436 instances in total), followed by additive (106 estimated instances), adversative (62 instances), temporal (45 instances) and causal (6 instances) conjunction. The predominance of conditional conjunction is not surprising, given what is already known about legislative genres from sources such as Biel (2014). In the case of the very frequent additive cohesive conjunctions ‘and’ and ‘or’, translations are extremely predictable: all instances of ‘and’ become และ /lae/ and all instances of ‘or’ become หรือ /rue/. By way of comparison, and despite the fact that it is a lot less common, there is more variety in the translation of the additive conjunctive adjunct ‘furthermore’ where there are three translations for only four instances. Translations of the most frequent conditional conjunction ‘if’ are also fairly predictable, with three different translations for 293 estimated instances of cohesive ‘if’. As was the case with ‘and’ and ‘or’, translations of ‘if’ are generally morphologically very simple: ถ้า /tha/ or หาก /hak/, or a combination of the two, ถ้าหาก /tha+/hak/. Similar observations can be made about the translation of ‘but’, where there is one translation, แต่ /tae/, for 17 instances. Translations of the most common temporal form, ‘after’, buck the trend. They are less predictable (six different translations for 35 instances) and generally consist of at least two morphemes. Few generalizations can be made about causal conjunction (excluding conditional conjunction) in the parallel corpus, given sparse data, but it appears that translations here are consistent with the general observation that, for all categories, morphologically complex conjunctions/conjunctive adjuncts in English tend to be translated by morphologically complex conjunctions/conjunctive adjuncts in Thai. We note also that there are just two cases of shifts between different types of conjunction—temporal to conditional in Example 6.9 and conditional to an adversative in Example 6.23—so that our source and target texts are generally semantically convergent as well as morphologically convergent. Finally, in just nine out of 655 cases of cohesive conjunction do we note omissions. In general we can thus say that patterns of cohesive conjunction in our source texts are reflected in our target texts.

Our methodology so far shows us only what happens as a reflex to source text features however, and is thus very focused on source-target relations, or ‘equivalence’ relations,

as Biel (2014) puts it. It does not tell us whether the patterns we see in our target texts simply replicate those of the source text, or whether they would be normal for the target language anyway. In other words, they do not tell us anything about the ‘textual fit’ (Biel *ibid.*) of our translations with non-translated Thai legislative texts in general. This is what the next section is designed to do.

### ***Cohesive Conjunction in Legal Thai Monolingual and Thai Translated Texts***

In this section, because we cannot obtain a list of Thai cohesive conjunctions/conjunctive adjuncts directly from Wordlist in Sketch Engine, we compare the frequency of all Thai conjunctions/conjunctive adjuncts obtained from the investigation so far in this chapter (that is, from the Thai translations of English conjunctions/conjunctive adjuncts) with their frequency in the Thai monolingual corpus. We also calculate the log likelihood (LL) score to see if there is any significant difference in their distribution in the two Thai corpora.<sup>30</sup>

We examine the distribution in both Thai corpora of instances of the five types of cohesive conjunction addressed above, and present results according to these types in the following.

#### **Additive Cohesive Conjunction**

We notice that the Thai monolingual corpus makes more use of basic conjunctions, namely และ /lae/ (and) and หรือ /rue/ (or) than the Thai translation corpus. We find the frequencies of two Thai conjunctions, และ /lae/ (and) and หรือ /rue/ (or) in the Thai monolingual corpus are significantly higher than those in the Thai translation corpus with the LL scores at 307.58 and 432.34 respectively.

On the other hand, the more complex conjunctions in the Thai monolingual corpus are either less frequent or similar to those in the Thai translation corpus. The frequency of the Thai conjunction นอกจากนี้ /nok/+/chak/+/ni/ (additionally, moreover, furthermore) in the Thai monolingual corpus is significantly lower than that in the Thai translation corpus with the LL score at 29.38 whereas the frequencies of the other three conjunctions, นอกเหนือจากนี้ /nok/+/nuea/+/chak/+/ni/ (additionally, moreover,

---

<sup>30</sup> As already indicated in Chapter 6, if the log likelihood is greater than 6.63, the probability of the result happening by chance is less than 1%. If the log likelihood is 3.84 or more, the probability of it happening by chance is less than 5%.

furthermore), ยิ่งไปกว่านั้น /ying/+/pai/+/kwa/+/nan/ (additionally, moreover, furthermore), and ยิ่งกว่านั้น /ying/+/kwa/+/nan/ (additionally, moreover, furthermore) are not significantly different between these two corpora. Having said that, the number of the Thai complex additive conjunctions remains low in *both* corpora. There are only 8 instances of นอกจากนี้ /nok/+/chak/+/ni/, and one instance each of นอกเหนือจากนี้ /nok/+/nuea/+/chak/+/ni/, ยิ่งไปกว่านั้น /ying/+/pai/+/kwa/+/nan/) and ยิ่งกว่านั้น /ying/+/kwa/+/nan/ in the Thai translation corpus, and none is found in the Thai monolingual corpus, as summarized in Table 6.7.

Thai additive conjunction	Thai monolingual corpus (1,173,485 words)	Thai translation corpus (222,556 words)	LL Score
	(Estimated) Number of cohesive conjunctive uses (frequency per million)	(Estimated) Number of cohesive conjunctive uses (frequency per million)	
และ /lae/ (and)	1,990 (1,685.80) <sup>31</sup>	90 (404.39) <sup>32</sup>	307.58
หรือ/rue/ (or)	1,810 (1,542) <sup>33</sup>	30 (134.79) <sup>34</sup>	432.34
นอกจากนี้ /nok/+chak/+ni/ (additionally, moreover, furthermore)	0 (0)	8 (35.35)	29.38
นอกเหนือจากนี้ /nok/+nuea/+chak/ +/ni/(additionally, moreover, furthermore)	0 (0)	1 (4.42)	3.67
ยิ่งไปกว่านั้น /ying/+pai/+kwa/+/ nan/(additionally, moreover, furthermore)	0 (0)	1 (4.42)	3.67
ยิ่งกว่านั้น /ying/+kwa/+nan/( additionally, moreover, furthermore)	0 (0)	1 (4.42)	3.67

**Table 6.7: Additive conjunction in the Thai monolingual and translation corpora**

#### Adversative Cohesive Conjunction

In the case of adversative conjunction (Table 6.8), we notice that morphologically simple conjunctions are more common in the Thai monolingual corpus while more complex conjunctions/conjunctive adjuncts are more frequent in the Thai translation corpus. The conjunction แต่ /tae/ (but) is significantly more frequent in the Thai monolingual corpus with an LL score of 133.59, while the three conjunctions อย่างไรก็ตาม /yang/+rai/+ko/+tam/, อย่างไรก็ตามก็ดี /yang/+rai/+ko/+di/ and แม้กระนั้น /mae/+kra/+nan/, all

<sup>31</sup> The actual frequency of และ /lae/ (in all its functions) in the Thai monolingual corpus is 22,887.

<sup>32</sup> The actual frequency of และ /lae/ (in all its functions) in the Thai translation corpus is 6,308.

<sup>33</sup> The actual frequency of หรือ/rue/ (in all its functions) in the Thai monolingual corpus is 26,807.

<sup>34</sup> The actual frequency of หรือ/rue/ (in all its functions) in the Thai translation corpus is 2,948.

of which can be glossed as ‘although/even though/nevertheless/nonetheless’, are significantly more frequent in the Thai translation corpus with LL scores of 71.76, 5.47 and 7.34 respectively. Finally, the frequencies of **แม้ว่า** /mae/+/wa/ (although/even though/nevertheless/nonetheless) in the two corpora are not significantly different with an LL score of 0.05. That said, as was the case with additive conjunction above, we again notice the infrequent use of complex conjunctions in both Thai corpora.

Thai adversative conjunction	Thai monolingual corpus (1,173,485 words)	Thai translation corpus (222,556 words)	LL Scores
	(Estimated) No. of cohesive conjunctive uses (frequency per million)	(Estimated) No. of cohesive conjunctive uses (frequency per million)	
<b>แต่</b> /tae/ (but)	650 (553.90) <sup>35</sup>	16 (72.56) <sup>36</sup>	133.59
<b>อย่างไรก็ตาม</b> /yang/+/rai/+/ko/+/tam/ (although, even though, nevertheless, nonetheless)	10 (8.24)	31 (139.29)	71.76
<b>แม้ว่า</b> /mae/+/wa/ (although, even though, nevertheless, nonetheless)	46 (39.70)	8 (35.35)	0.05
<b>อย่างไรก็ดี</b> /yang/+/rai/+/ko/+/di/(although, even though, nevertheless, nonetheless)	11 (9.06)	7 (30.93)	5.47
<b>แม้กระนั้น</b> /mae/+/kra/+/nan/(although, even though, nevertheless, nonetheless)	0 (0)	2(8.84)	7.34

**Table 6.8: Adversative conjunction in the Thai monolingual and translation corpora**

#### Temporal Cohesive Conjunction

As we can see from Table 6.9, a simple conjunction, namely **ก่อน** /kon/ (before) is used significantly more in the Thai monolingual corpus than in the Thai translation corpus (with an LL scores of 10.65), while five more complex conjunctions/conjunctive

<sup>35</sup> The actual frequency of **แต่** /tae/ (in all its functions) in the Thai monolingual corpus is 5,854.

<sup>36</sup> The actual frequency of **แต่** /tae/ (in all its functions) in the Thai translation corpus is 323.



adjuncts, namely หลังจาก /lung/+/chak/ (after), ภายหลัง /pai/+/lung/ (after), ถัดจาก /that/+/chak/ (after), นับแต่ /nup/+/tae/ (since), จากนั้น /chak/+/nan/ (then/after) are much preferred in the Thai Translation corpus. The frequencies of the two remaining complex conjunctions, นับจาก /nup/+/chak/ (since) and ในขณะที่ /nai/+/kha/+/na/+/ti/ (while), in the Thai monolingual corpus are not significantly different from those in the Thai translation. Overall, the number of instances of temporal cohesive conjunction is relatively low in both Thai corpora.

Thai temporal conjunction	Thai monolingual corpus (1,173,485 words)	Thai translation corpus (222,556 words)	LL Scores
	(Estimated) Number of cohesive conjunctive uses (frequency per million)	(Estimated) Number of cohesive conjunctive uses (frequency per million)	
หลังจาก /lung/+/chak/ (after)	15 (12.78)	44 (197.7)	99
ภายหลัง /pai/+/lung/ (after)	6 (5.11)	10 (44.93)	17.64
ถัดจาก /that/+/chak/ (after)	0 (0)	2 (8.98)	7.34
นับจาก /nup/+/chak/ (since)	0 (0)	1 (4.49)	3.67
นับแต่ /nup/+/tae/ (since)	0 (0)	2 (8.98)	7.34
ก่อน /kon/ (before)	100 (85.21) <sup>37</sup>	6 (26.95)	10.65
ในขณะที่ /nai/+/kha/+/na/+/ti/ (while)	21 (17.89)	6 (26.95)	0.72
จากนั้น /chak/+/nan/ (then/after)	2 (1.70)	6 (26.95)	13.73

**Table 6.9: Temporal conjunction in the Thai monolingual and translation corpora**

#### Causal Cohesive Conjunction

The frequencies of all Thai causal cohesive conjunctions (excluding conditional ones) in both corpora are very low, all at under 100 occurrences per million words (Table 6.10). In addition, there are no significant differences between the two corpora in the frequencies of causal cohesive conjunctions/conjunctive adjuncts.

<sup>37</sup> The actual frequency of ก่อน /kon/ (in all its functions) in the Thai monolingual corpus is 2,029.

Thai causal conjunction	Thai monolingual corpus (1,173,485 words)	Thai translation corpus (222,556 words)	LL Scores
	(Estimated) Number of cohesive conjunctive uses (frequency per million)	(Estimated) Number of cohesive conjunctive uses (frequency per million)	
เพราะว่า /pro/+wa/ (because)	0 (0)	1 (4.49)	3.67
เพราะ /pro/ (because)	22 (18.75)	1 (4.49)	3.09
ด้วยเหตุนี้ /duai+/hed+/ni/ (accordingly)	1 (0.82)	1 (4.49)	1.25
ในการนี้ /nai+/kan+/ni/ (accordingly)	116 (95.58)	14 (62.90)	2.87
ฉะนั้น /cha+/nan/ (therefore)	11 (9.37)	1 (4.49)	0.61

**Table 6.10: Causal conjunction in the Thai monolingual and translation corpora**

#### Conditional Cohesive Conjunction

As previously mentioned, we consider the category of conditional cohesive conjunction separately due to its special role in legal texts. As Table 6.11 shows, we find that the Thai conditional cohesive conjunction ถ้า /tha/ (if) is the most frequently used conjunction in the Thai monolingual corpus (with a frequency of 2,641.70 per million words). We also find high frequencies for other common Thai conditional cohesive conjunctions, namely หาก /hak/ (if), เมื่อ /muea/ (when) and เว้นแต่ /wen+/tae/ (unless) in both corpora.

Consistent with the pattern already established in other categories, we see that the morphologically simple conditional conjunctions, namely ถ้า /tha/, เมื่อ /muea/ and เว้นแต่ /wen+/tae/, are more frequent in the Thai monolingual corpus than in the Thai translation corpus, while the more complex ones are more common in the Thai translation corpus.

Thai conditional conjunction	Thai monolingual corpus (1,173,485 words)	Thai translation corpus (222,556 words)	LL Scores
	(Estimated) Number of cohesive conjunctive uses (frequency per million)	(Estimated) Number of cohesive conjunctive uses (frequency per million)	
ถ้า /tha/ (if)	3,100 (2,641.70) <sup>38</sup>	80 (359.46) <sup>39</sup>	623.31
หาก /hak/ (if)	599 (510.45) <sup>40</sup>	124 (557.16) <sup>41</sup>	0.77
ถ้าหาก /tha+/hak/ (if)	12 (10.22)	14 (62.90)	19.69
โดยมีเงื่อนไขว่า /doi+/mi+/nguean+/khai+/wa/ (under the condition that)	8 (6.59)	11 (49.42)	17.31
ภายใต้เงื่อนไข ว่า/phai+/tai+/nguean+/khai+/wa/ (under the condition that)	0 (0)	2 (8.98)	7.34
เมื่อ /muea/ (when)	970 (826.60) <sup>42</sup>	151 (679.37) <sup>43</sup>	5.34
เว้นแต่ /wen+/tae/ (unless)	350 (298.3)	41 (184.22)	9.67
ก็ต่อเมื่อ /ko+/to+/muea/ (only if)	5 (4.26)	2 (8.98)	0.71
ตราบเท่าที่ /trap+/tao+/ti/ (as long as)	6 (5.11)	3 (13.47)	1.64
ทั้งนี้ /tang+/ni/ (so as to)	260 (221.5)	40 (179.73)	1.60
เว้นไว้แต่ /wen+/wai+/tae/ (unless)	7 (5.77)	2 (8.98)	0.24
ยกเว้นแต่ /yok+/wen+/tae/ (unless/except)	1 (0.85)	5 (22.46)	13.30
เว้นเสียแต่ /wen+/sia+/tae/ (unless/except)	0 (0)	2 (8.98)	7.34
ยกเว้น /yok+/wen/ (except)	0 (0)	4 (17.97)	14.69
นอกจาก /nok+/chak/ (except)	9 (8)	4 (17.97)	1.77
นอกเสียจาก /nok+/sia+/chak/ (except)	0 (0)	1 (4.49)	3.67
หากไม่... /hak/.../mai/ (if...not)	0 (0)	30 (134.79)	110.17
ตราบใดที่..... ไม่ /trap+/dai+/tee/.../mai/ (as long as...not)	7 (5.77)	1 (4.49)	0.08

**Table 6.11 Conditional conjunction in the Thai monolingual and translation corpora**

<sup>38</sup> The actual frequency of the form ถ้า /tha/ (in all its functions) in the Thai monolingual corpus is 3,457.

<sup>39</sup> The actual frequency of the form ถ้า /tha/ (in all its functions) in the Thai translation corpus is 154.

<sup>40</sup> The actual frequency of the form หาก /hak/ (in all its functions) in the Thai monolingual corpus is 809.

<sup>41</sup> The actual frequency of the form หาก /hak/ (in all its functions) in the Thai translation corpus is 274.

<sup>42</sup> The actual frequency of the form เมื่อ /muea/ (in all its functions) in the Thai monolingual corpus is 3,089.

<sup>43</sup> The actual frequency of the form เมื่อ /muea/ (in all its functions) in the Thai translation corpus is 315.

### *Cohesive Conjunctions in English vs Thai Translated Texts vs Thai Non-Translated Texts*

Thus far we have seen that patterns of cohesive conjunction in our target texts follow closely those observed in the source texts from a formal, a semantic and a distributional point of view. We have also seen that certain basic cohesive conjunctions are significantly more frequent in our monolingual Thai corpus than in translated Thai. These observations suggest that cohesive conjunction is a resource that is used to a far greater extent in monolingual Thai legislation than in either English source texts or Thai translated legislation. Monolingual Thai thus appears to rely more on coordinating conjunctions, and thus parataxis, than either English source texts or Thai translations. To further probe this idea, we conduct one final comparison, between the distributions in our English source texts and our Thai monolingual corpus of the most common, basic conjunctions, for which translations appear to be fairly stable in our parallel corpus, and where we can thus assume a valid *tertium comparationis*: hence we compare the distributions of ‘and’, ‘or’, ‘but’ and ‘if’, and their translation equivalents และ /lae/, หรือ /rue/, แต่ /tae/ and, in the case of ‘if’, ถ้า /tha/, หาก /hak/ and ถ้าหาก /tha+/hak/, in the two bilingually comparable corpora. As the latter three Thai conjunctions account for all the translations of ‘if’ in our sample, we combine their frequencies to estimate the frequency of conjunctions equivalent to ‘if’ in the Thai monolingual corpus. We use log likelihood scores to indicate whether differences between the distributions of equivalent words in English and (monolingual) Thai are significant. Results are shown in Table 6.12.

English cohesive conjunction (est. frequency per million)	(Estimated) Frequency in corpus (size: 172,739)	LL Scores	(Estimated) Frequency in corpus (size: 1,173,485)	Thai non-translated cohesive conjunction (frequency per million)
and (521)	90	174.84	1,990	และ /lae/ (1,685.80)
or (57.89)	10	432.63	1,810	หรือ /rue/ (1,542)
but (98.41)	17	90.01	650	แต่ /tae/ (553.90)
if (1,696.20)	293	38.23	3,711	ถ้า /tha/ (if), หาก /hak/ (if), ถ้าหาก /tha+/hak/

**Table 6.12: Comparison of frequency of cohesive conjunctions in English and Monolingual Thai corpora**

Overall, it can be seen that the Thai monolingual corpus has significantly higher frequencies of cohesive conjunctions equivalent to ‘and’, ‘or’, ‘but’ and ‘if’ compared to the comparable English corpus.

## **Discussion**

We find that, in general, there is a low incidence of cohesive conjunction in English, and the Thai translated conjunctions very much follow their English counterparts, formally, semantically and distributionally. On very rare occasions, there are omissions and shifts in semantic relations in the Thai translations. A possible reason for tight adherence to source-text patterns is the normative nature of the English source texts and the non-authoritative application of the translated texts in the Thai context. The translations of these legal texts were officially made by the professionals assigned by the relevant authorities, such as the Ministry of Foreign Affairs, and they are mostly used internally by such authorities. As already mentioned, the Thai translations are not authoritative; nevertheless, since the English language does not have any status in Thailand, many users have to rely on Thai translated versions of treaties. It can be assumed that translators are likely to preserve as many details as possible of the source text in their translations, so as not to create any confusion regarding the normative intent of the texts.

In the second part, we compare the distribution of the Thai cohesive conjunctions in the Thai monolingual corpus and Thai translation corpus. Based on the 42 Thai conjunctions, we find that the frequencies of 19 rare and complex Thai conjunctions are not significantly different between these two corpora; some of these rare conjunctions occur only once, however, and it is hard to generalize on this basis. There are 7 Thai conjunctions that are more frequent in the Thai monolingual corpus than in the Thai translation corpus, and 16 Thai conjunctions that are more frequent in the Thai translation corpus than in the Thai monolingual corpus. It is interesting that the 7 conjunctions that are very frequent in the Thai monolingual corpus are all simple and common conjunctions while the 16 conjunctions that are used much more in legal Thai translation corpus are more complex conjunctions. Given the preponderance of basic conjunctions in the Thai monolingual corpus, it is possible that legislative Thai is, on the whole, more explicit when it comes to additive and basic adversative cohesive conjunction than either our English source texts, but a serious study of this would require a separate contrastive textological study, that is beyond the scope of our

research. For now, we conclude simply that when it comes to cohesive conjunction, our study finds no evidence of explicitation.

Regarding unique items, we find one Thai conjunction <sup>๓๓๓</sup> /tang/+/ni/ which does not have a straightforward equivalent in English and could be considered a unique item based on Tirkkonen-Condit's (2004) definition. We find that the frequency of <sup>๓๓๓</sup> /tang/+/ni/ in the Thai translation corpus is just slightly lower than that in Thai monolingual but the difference is not significant (with LL score 1.60). We thus cannot consider this item to be underrepresented in Thai translation, and cannot consider cohesive conjunction overall as a particularly fruitful avenue for the investigation of unique items.

As observed in the first part of the study reported on in this chapter, the Thai translated cohesive conjunctions strictly follow their English counterparts. We then consider if this might fit into the concept of 'source language shining through' proposed by Teich (2003) where translations are oriented more towards their source texts, which, in turn, makes them different from comparable texts in the same language. In other words, source language shining through lessens the textual fit of the translations with other texts of the same genre in the target language. Nevertheless, based on findings in the third part, although the distribution of Thai translated cohesive conjunctions is much closer to that of English cohesive conjunctions than that of the Thai monolingual corpus, the distributions of certain conjunctions in the Thai translation corpus are also significantly different from those in the source text. In an attempt to identify the distinguishing features of Thai legislative translation based on the investigation of cohesive conjunctions in our corpora, at this stage, we can roughly conclude that the use of cohesive conjunctions in the Thai translated texts under investigation is more like that of their English source texts than that of their Thai comparable texts. The evidence we have also suggests that the cohesive conjunctions in the Thai translations exhibit specific patterns that differ from those in the English source text and non-translated comparable texts in the same language.

## Conclusions

In this chapter, we investigate the use of cohesive conjunction in our English source texts, our Thai translations, and comparable monolingual texts in Thai. We find evidence of the source language shining through the translated texts as translated

cohesive conjunctions strictly follow their English counterparts. At the same time we find a low level of textual fit between Thai translations and monolingual, non-translated Thai. This low level of textual fit is possibly due to radically different ways of using paratactic conjunctions in English and Thai, a subject that would benefit from more thorough research.

## **Chapter 7: Conclusions**

This chapter summarizes the issues discussed in the preceding chapters. The main theories and methodology employed are briefly explained while the results of the investigation are mentioned. The research contribution is illustrated, followed by research implications and limitations. Finally, areas for possible further study are proposed.

### **Theoretical Framework and Methodology**

In this thesis, we conduct inter-linguistic and intra-linguistic comparisons based on two self-built corpora: an English-Thai parallel corpus of legislative texts and a monolingual corpus of non-translated legislative texts in Thai. We investigate three linguistic features: passive, deontic modality and cohesive conjunction. We investigate whether there are manifestations of certain translation features, namely ‘explicitation’ (Becher 2011), ‘source language shining through’ (Teich 2003) and the underrepresentation of ‘unique items’ (Tirkkonen-Condit 2002) in English-Thai translations of these three linguistic items.

The methodology we employ in this study mainly relies on Biel’s (2014) study where the concepts of ‘equivalence’ and ‘textual fit’ are applied using corpora. These concepts are based on Chesterman’s (2004) potential S-Universals and T-Universals respectively. While the ‘equivalence’ or S-Universals attempt to capture generalizations about differences between translations and their source texts, the ‘textual fit’ or T-Universals focus on differences between translation and comparable non-translated texts (Chesterman 2004, 2010).

Like Biel’s (2014), this study is predominantly data-driven in that findings from our corpora are used as the starting point of the analysis. This thesis draws on two self-built corpora: a bilingual parallel corpus consisting of English legislative texts (172,739 words) aligned with their Thai translations (222,556 words); and a monolingual corpus consisting of legislative texts originally written in Thai (1,173,485 words). Sketch Engine’s Concordance tool is employed for the searches to find the relationship between translated legislation and source texts, and translated and non-translated legislative texts in the target language.



It is obvious that the explicitation hypothesis can be tested as a potential S-universal/equivalence relation, by comparing certain features in the English source texts and their Thai translations, while the underrepresentation of target language specific items (Tirkkonen-Condit 2002) belongs to the class of potential T-universals/textual fit, which can be tested by comparing the Thai translated texts with non-translated Thai. However, when we examine the evidence of source language shining through, on some occasions we compare the linguistic items not only between the source and target texts but also across the three sets of texts. The following section briefly summarizes the results of our investigation of Thai translations of passive voice, deontic modality and cohesive conjunction.

### **Recapitulation of Research**

In this section the findings from our investigations of passive, deontic modality and cohesive conjunction are presented based on two main relations: the relation between English source texts and Thai translated texts, and the relation between Thai translated texts and Thai non-translated comparable texts.

Regarding passive, when we examine the relation between source and target texts, our data show that the translation of English passive voice into Thai provides only weak evidence to support the explicitation hypothesis. Considering that passive is not favoured in Thai, we expected to see instances of English passive voices translated into Thai active voice, perhaps with the concomitant explicitation of agents. However, we find that most English passives are translated into Thai using passive voice and that in the instances where active voice is used in Thai, we rarely find any explicitation involved, as explicit agents are rarely added in Thai.

In addition, in terms of textual fit, the monolingual comparable data disconfirms the unique item hypothesis (Tirkkonen-Condit 2002). The passive markers *ถูก* /thuk/ and *ได้รับ* /dai/+/rup/ which are considered unique items in Thai passive are overrepresented rather than underrepresented in translated data compared to non-translated data. The overrepresentation of these two passive markers in Thai translated texts (roughly over 40% higher than that in the monolingual text) might be influenced by the passive-laden source texts. This therefore, could be considered that the use of passive in the source text shines through that in Thai translation.

With respect to Thai translated deontic modal verbs, our data do not provide evidence to confirm the explicitation hypothesis. The majority of English deontic modal verbs are translated into somewhat predictable patterns in Thai. Approximately one third of Thai translated modal verbs are translated in a way that syntactically diverges from their English counterparts. In these cases, the English modal verbs are translated into Thai simple declarative sentences and Thai causative structures rather than Thai modal verbs; however, the semantic element of obligation remains unaltered when we examine the modal strength of these two patterns. However, we notice that when the English modal verbs are translated into Thai simple declarative sentences, this might constitute cases of implicitation as posited by Becher (2010) where information that was verbalized in the source text is not verbalized in the target text.

Investigating textual fit, we find that based on the monolingual comparable data, the only Thai modal verb *จะต้อง* /cha+/tong/ which is considered unique to Thai is overrepresented in the Thai translation, so our finding here disconfirms the unique item hypothesis. In addition, we find that Thai modal verbs are much more frequent in translated than in non-translated Thai but still less frequent than in the English corpus. This can be due to the interference of the source text or source language shining through where the use of modal verbs in Thai translation is influenced by that in the English source texts. However, we note that the distribution of modal verbs in Thai translation lies halfway between the norms of English and Thai. Therefore, we speculate that the translators might be steering a middle course in their use of deontic modal verbs.

In relation to the investigation of cohesive conjunction, based on the findings from the parallel corpus, we might report that we do not find any evidence of explicitation in the Thai translations of the English cohesive conjunctions. Thai translated conjunctions very much follow their English counterparts, and mostly they maintain the same semantic content as their English counterparts.

In terms of textual fit, regarding the unique items, we do not find strong evidence to support Tirkkonen-Condit's (2002) hypothesis, so this investigation cannot confirm the unique item hypothesis. We then explore the distributions of the cohesive conjunctions across three texts. We find that the distribution of cohesive conjunctions in Thai translated texts is different from that in English source texts and the Thai monolingual texts. Unlike our finding in the deontic modality section, the distribution of selected

conjunctions in the Thai translation corpus is not seen to lie half way between norms for the two languages. Rather, the distribution of Thai translated cohesive conjunctions is much closer to that of English cohesive conjunctions than to that of our Thai monolingual corpus. Nevertheless, the distributions of selected conjunctions in the Thai translation corpus are also still significantly different from those in the source text. We might explain the pattern used in translation as being partly influenced by the source language shining through.

In conclusion, based on our findings, this research cannot support either the explicitation and or unique item hypothesis. We also see that the co-occurrence of source language shining through translation and the process of normalization make Thai legislative translation different from English source texts and Thai comparable texts.

### **Contribution of the Research**

This research is a first attempt to use corpus techniques to explore legislative translation from English into Thai. As previously mentioned, interest in legislative translation has grown substantially over recent decades, with corpus-based approaches contributing to our understanding of the relationship between translated legislation and source texts, on the one hand, and translated and non-translated legislative texts in the target language, on the other. To date, nevertheless, most studies have been conducted on European languages.

We believe that the thesis has contributed to the literature on translation studies. Our exploration involved investigation of how English passive voice, deontic modality and cohesive conjunction are translated into Thai in the light of the translation features. We have provided more empirical evidence on this topic than previous discussions, which were not based on corpora. The fact that the findings reveal the specific patterns used in Thai legal translated texts does not have an impact only on the area of legal translation but also on translation studies as a whole, because translation studies still require new perspectives especially from under-researched languages like Thai. We believe also that our work can contribute to contrastive linguistic investigations in the English-Thai language pair, where to date there has been a dearth of research. In addition, our findings are also potentially of interest to statistical, including neural machine translation, where an understanding of translation features is useful in explaining

differences in the performance of statistical machine translation engines trained on different data sets (Lembersky et al 2012, 2013).

With regard to the contribution of the thesis to Thai translation culture in particular, this thesis breaks new ground in studying Thai linguistic features used in Thai legal texts, translated and original. The fact that the findings of the thesis reveal specific patterns that were not noticed before both in the Thai legal translated texts and Thai legal non-translated texts contributes not only to the area of Thai legal translation studies, but also of Thai legal drafting in general. The study contributes to knowledge of practice because it reveals previously hidden aspects of the process and product of Thai legal translation and this could be of interest to producers of legislation.

In addition, this thesis also contributes to Thai corpus linguistics. To our knowledge, this is the first attempt to construct a Thai-English legal parallel corpus and a Thai legal monolingual corpus from scratch. We believe that these two corpora could be a prototype and will no doubt be used in follow-up studies, in translation and contrastive linguistics.

Finally, given the methodological innovation required for this thesis, we are in a position to give advice to future researchers who may wish to pursue corpus-based studies of Thai. This advice is encapsulated in Appendix D.

### **Limitations of the Research and Avenues for Further Study**

This thesis is subject to two main limitations. The first limitation is concerned with the scarcity of literature on the Thai side. We find that the literature on the Thai side is very limited which makes our study leave many questions unanswered. For example, when we analyze data from our investigation of deontic modality, we find that the study is limited by the fact that Thai causatives and simple declarative structures that indicate obligation or permission have not yet been systematically studied, and that full explanations have yet to be arrived at for the distribution of modal forms in Thai. Such explanations would benefit from reversing the direction of the analysis, and proceeding from Thai into English. Second, a corpus tool which can fully support Thai still does not exist. Although Sketch Engine works quite well with Thai, there are still some limitations when operating search queries. There are many functions which do not support Thai, making it difficult to search for instances of structures that are not signalled by specific lexical items. For example, we cannot investigate how the use of

*covert* passive differs between translated and non-translated Thai because we cannot yet search for this pattern on the Thai side.

Despite these shortcomings, this research still reveals specific patterns in legislative texts translated into Thai which might be useful in translation studies. In addition, having already created the corpus resources used in this study, we are in a position to widen our investigation to consider other manifestations of other related linguistic features in legal translation from English to Thai, for example we might consider the effects of sub-genres in future research.

## References

- Abdelali A., Guzman, F., Sajjad, H., and Vogel, S. (2014): The AMARA Corpus: Building Parallel Language Resources for the Educational Domain, in *LREC 2014*. Available at: <http://alt.qcri.org/~guzmanhe/papers/LREC2014-Abdelali.pdf> (Accessed 10 June 2018)
- Adams, K. and Kaye, A. (2006) 'Revisiting the ambiguity of "and" and "or" in legal drafting', *St. John's Law Review*, 80(4), pp. 1167-1198.
- Aijmer, K. (2016) 'Modality and mood in functional linguistic approaches', in Nuyts, J. and Van Der Auwera, J. (eds.) *The Oxford Handbook of Mood and Modality*. 1<sup>st</sup> ed. Oxford: Oxford University Press, pp. 495-513.
- Ajulo, S.B. (1997) 'Myth and reality of law, language and international organization in Africa: The case of African economic community', *Journal of African Law*, 41(01), pp. 27-42.
- Alcaraz, E. and Hughes, B. (2002) *Legal Translation Explained*. Manchester: St. Jerome.
- Anuman, R. (1961) *The Thai Language*. (Thailand Cultural Series No. 17). Bangkok: The National Culture Institute.
- Aroonmanakun, W. (2007) 'Creating the Thai national corpus', *Manusaya*. Special Issue. 13, pp. 4-17.
- Aroonmanakun, W., Tansiri, K. and Nittayanuparp, P. (2009) 'Thai national corpus: A progress report', *Proceedings of the 7th Workshop on Asian Language Resources, ACL-IJCNLP 2009*. Suntec, Singapore, 6-7 August 2009, pp. 153-160. Available at: <http://pioneer.chula.ac.th/~awirote/ling/TNC2009.pdf> (Accessed 10 June 2018)
- Aust, A. (2003) *Modern Treaty Law and Practice*. Cambridge: Cambridge University Press.

- Baker, M. (1993) 'Corpus linguistics and translation studies: Implications and applications', in Baker, M., Francis, G. and Tognini-Bonelli, E. (eds.) *Text and Technology: In Honour of John Sinclair*. Amsterdam: John Benjamins, pp. 17-45.
- Baker, M. (1995) 'Corpora in translation studies: An overview and some suggestions for future research', *Target* 7(2), pp. 223-245.
- Baker, M. (1996) 'Corpus-based translation studies: The challenges that lie ahead', in Somers, H. (ed.) *Terminology, LSP and Translation: Studies in Languages Engineering, in Honor of Juan C. Sager*. Amsterdam: John Benjamins, pp. 175-186.
- Baker, M. (1992) *In Other Word*. 1<sup>st</sup> ed. London; New York: Routledge.
- Baker, M. (2018) *In Other Words*. 3<sup>rd</sup> ed. Oxon; New York: Routledge.
- Balaza, M. (2014) 'Legal translation as an interspace of language and law', in Boldea, I. (ed.) *Studies on Literature, Discourse and Multicultural Dialogue*. Tirgu Mures: Arhipelag XXI Publishing House, pp. 357-364.
- Barlow, M. (2003) *ParaConc*, Available at: [www.athel.com](http://www.athel.com).
- Beaupré, M. (1986) *Interpreting Bilingual Legislation*. Toronto: Carswell.
- Becher, V. (2010a) 'Towards a more rigorous treatment of the explicitation hypothesis in translation studies', *trans-kom* 3(1), pp. 1-25.
- Becher, V. (2010b) 'Abandoning the notion of 'translation-inherent' explicitation: against a dogma of translation studies', *Across Languages and Cultures* 11(1), pp. 1-28.
- Becher, V. (2011) *Explicitation and Implication in Translation: A Corpus-Based Study of English-German and German-English Translation of Business Texts*. PhD thesis. University of Hamburg.
- Bernardini, S. and Zanettin, F. (2004) 'When is a universal not a universal? Some limits of current corpus-based methodologies for the investigation of translation universal' in Mauranen, A. and Kujamäki, P. (eds.) *Translation Universals: Do They Exist?* Amsterdam; Philadelphia: John Benjamins, pp. 51-64.

Bhatia, V. (1994) 'Cognitive structuring in legislative provisions', in Gibbons, J (ed.) *Language in the Law*. London: Longman, pp. 1-16.

Biel, Ł. (2010) Corpus-based studies of legal language for translation purposes: Methodological and practical potential', in Heine, C. and Engberg, J. (eds.) *Reconceptualizing LSP. Online proceedings of the XVII European LSP Symposium 2009*. Aarhus 2010. pp. 1-15.

Biel, Ł. (2014) *Lost in the Eurofog: The Textual Fit of Translated Law*. Frankfurt am Main: Peter Lang.

Biel, Ł. and Engberg, J. (2013) 'Research models and methods in legal translation', in Biel, Ł. and Engberg, J. (eds.) *Linguistica Antverpiensia*, Special Issue on Research Models and Methods in Legal Translation ,12, pp.1-11.

Bisang, W. (2017) 'Grammaticalization', *Oxford Research Encyclopaedia of Linguistics*. Oxford: OUP. Published online March 2017 Available at: <http://dx.doi.org/10.1093/acrefore/9780199384655.013.103> (Accessed 1 June 2018)

Bloor, T. and Bloor, M. (2004) *The Functional Analysis of English. A Hallidayan Approach*. 2<sup>nd</sup> ed. London: Oxford University Press.

Blum-Kulka, S. (1986) 'Shifts of cohesion and coherence in translation', in House, J. and Blum-Kulka, S. (eds.) *Interlingua and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies*. Tübingen: Gunter Narr, pp. 17-35.

Boonchalermwipas, S. (2009) *Thai Legal History*. Bangkok: Winyuchon Publishing.

Bowker, L. and Pearson, J. (2002) *Working with Specialized Language: A Practical Guide to Using Corpora*. London: Routledge.

Brown, J. (1985) *From Ancient Thai to Modern Dialects*. Bangkok: Sivaphorn Limited Partnership.

Burukina, O. (2012) 'Legal language: A realm of contradictions', *Contemporary Readings in Law & Social Justice*, 4(2), pp. 708-723.



- Cabanellas, G. (2014) *The Legal Environment of Translation*. Oxon: Routledge.
- Cao, D. (2007) *Translating Law*. Clevedon: Multilingual Matters.
- Cao, D. (2010) 'Legal translation', in Coulthard, M. and Johnson, A. (eds.) *The Routledge Handbook of Forensic Linguistics*. Oxon: Routledge, pp.78-91.
- Cao, D. (2013) 'Legal translation studies', in Millán, C. and Bartrina, F. (eds.) *The Routledge Handbook of Translation Studies*. New York: Routledge.
- Cappelle, B. (2012) 'English is less rich in manner-of-motion verbs when translated from French', *Across Languages and Cultures* 13 (2), pp. 173–195.
- Cappelle, B. and Looock, R. (2017) 'Typological differences shining through: The case of phrasal verbs in translated English', in Sutter G.D., Lefer M. and Delaere, I. (eds.) *Empirical Translation Studies. New Theoretical and Methodological Traditions*. Berlin: Mouton de Gruyter, pp. 235-264.
- Carter, R. and McCarthy, M. (2006) *Cambridge Grammar of English. A Comprehensive Guide. Spoken and Written English Grammar and Usage*. Cambridge: Cambridge University Press.
- Catford, J.C. (1965) *A Linguistic Theory of Translation: An Essay in Applied Linguistics*. Oxford: Oxford University Press.
- Chanawangsa, W. (1986) *Cohesion in Thai*. PhD thesis. Georgetown University.
- Charoenpornasawat, P. (2003) SWATH Available at:  
<http://www.cs.cmu.edu/~paisarn/software.html>
- Chen, W. (2004) 'Investigating explicitation of conjunctions in translated Chinese: A corpus-based study', *Language Matters: Studies in the Languages of Africa*, 35(1), pp. 295-312.
- Cheong, H. (2006) 'Target Text Contraction in English-into-Korean Translations: A Contradiction of Presumed Translation Universals?', *Meta* 51(2), pp. 343–367.

Chesterman, A. (2004) 'Beyond the Particular', in Mauranen, A. and Kujamäki, P. (eds.) *Translation Universals: Do They Exist?* Amsterdam; Philadelphia: John Benjamins, pp. 33-49.

Chesterman, A. (2007) 'What is a unique item?' in Gambier, Y., Schlesinger, M. and Stolze, R. (eds.) *Doubts and Directions in Translation Studies*. Amsterdam; Philadelphia: John Benjamins, pp. 3-13.

Chesterman, A. (2010) 'Why study translation universals?' *Kiasm*, 1, pp. 38-48.

Christodouloupoulos, C. and Steedman, M. (2015) 'A massively parallel corpus: the Bible in 100 languages'. *Language Resources and Evaluation* 49(2), pp. 375-395. Available at <https://doi.org/10.1007/s10579-014-9287-y> (Accessed 10 June 2018)

*Copyright Act of Thailand* (1994). Available at: <https://www.ipthailand.go.th/th/dip-law-2/item/copyright-act-b-e-2537-1994.html>

Coulthard, M. and Johnson, A. (2007) *An Introduction to Forensic Linguistics: Language in Evidence*. London; New York: Routledge.

Dai, G. and Xiao, R. (2011) 'Source language shining through in translational language: A corpus-based study of Chinese translation of English passives', *Translation Quarterly*, 62, pp. 85–107.

De Sutter, G., Goethals, P., Leuschner, T. and Vandepitte, S. (2012) 'Towards methodologically more rigorous corpus-based translation studies'. *Across Languages and Cultures*, 13(2), pp. 137-143.

Diller, A. (2008) 'Resources for Thai language research', in Diller, A., et al (eds.) *The Tai-Kadai language*. London; New York: Routledge, pp.31-82.

Dullion, V. (2000) 'Du document à l'instrument : les fonctions de la traduction des lois'. *Actes du Colloque International "La traduction juridique. Histoire, théorie(s) et pratique,"* 17-19.2.2000, Université de Genève. Berne: ASTTI, 233-253. Available at : <http://www.tradulex.com/Actes2000/dullion.pdf> (Accessed 18 June 2018)

Durieux, C. (1991) 'Liberté et créativité en traduction technique', in Lederer, M. and Israel, F. (eds.), *La liberté en traduction*. Paris: Didier Erudition, pp. 169–179.

Englund Dimitrova, B. (2005) *Expertise and Explicitation in the Translation Process*. Amsterdam; Philadelphia: John Benjamins.

Eurolect Observatory (2017) Available at: <http://www.unint.eu/en/research/research-groups/39-higher-education/490-eurolect-observatory-interlingual-and-intralingual-analysis-of-legal-varieties-in-the-eu-setting.html>

Faber, D. and Hjort-Pedersen, M. (2009a) 'Manifestations of Interference Processes in Legal Translation', in Göpferich, S., Jakobsen, A. L. and Mees, I. (eds.) *Behind the Mind: Methods, Models and Results in Translation Process Research*. Copenhagen: Samfundslitteratur, pp. 107-124.

Faber, D. and Hjort-Pedersen, M. (2009b) 'Translation Preferences in Legal Translation: Lawyers and Professional Translators Compared', in Mees, I., Alves, F. and Göpferich, S. (eds.) *Methodology, Technology and Innovation in Translation Process Research: A Tribute to Arnt Lykke Jakobsen*. Copenhagen: Samfundslitteratur, pp. 339-357.

Faber, D. and Hjort-Pedersen, M. (2013) 'Expectancy and professional norms in legal translation: A study of explicitation and implicitation preferences', *Fachsprache* 35(1-2), pp. 42-62.

Garzone, G. (2000) 'Legal translation and functionalist approaches: A contradiction in terms?' *Legal Translation: History, Theory/ies and Practice. Proceeding of the International Colloquim*. Berne: University of Geneva, pp.349-356.

Garzone, G. (2000) 'Legal Translation and Functionalist Approaches: A Contradiction in Terms? Actes du Colloque International "La traduction juridique. Histoire, théorie(s) et pratique," 17-19.2.2000, Université de Genève. Berne: ASTTI, pp. 395-414. Available at: <http://www.tradulex.com/Actes2000/Garzone.pdf> (Accessed 16 April 2016)

Garzone, G. (2008) 'International commercial arbitration rules as translated/ re-written texts: an intercultural perspective', in Bhatia, V., Candlin C. and Allori, P. (eds.)

*Language, Culture and the Law: The Formulation of Legal Concepts across Systems and Culture*. Bern: Peter Lang, pp. 47-75.

Garzone, G. (2013) 'Variation in the use of modality in legislative texts: Focus on shall', *Journal of Pragmatics*, 57, pp. 68–81.

Gémar, J. (1995) *Traduire ou l'art d'interpréter, Langue, droit et société: éléments de jurilinguistique, tome2: Application-Traduire le texte juridique*, Saint-Nicolas (Québec): Presses de l'Université, Cowansville: Yvon Blais.

Gentzler, E. (2001) *Contemporary Translation Theories*. 2<sup>nd</sup> ed. Clevedon: Multilingual Matters.

Gibbons, J. (2003) *Forensic linguistics: An Introduction to Language in the Justice System*. Malden: Blackwell Publishing.

Gibbons, J. (2004) 'Taking legal language seriously', in Gibbons, J (ed.) *Language in the Law*. London: Longman, pp. 1-16.

Halliday, M. (1985). *An Introduction to Functional Grammar*. 1<sup>st</sup> ed. London: Edward Arnold.

Halliday, M. (1994) *An Introduction to Functional Grammar*. 2<sup>nd</sup> ed. London: Edward Arnold.

Halliday, M. and Hasan, R. (1976) *Cohesion in English*. London: Longman.

Halliday, M. and Matthiessen, C. (2004) *An Introduction to Functional Grammar*. 3rd ed. London: Arnold.

Hansen-Schirra, S., Neumann, S. and Steiner, E (2007) 'Cohesive explicitness and explicitation in an English-German translation corpus', *Languages in Contrast*, 7(2), pp.241-265.

Hareide, L. (2016) 'The translation of formal source-language lacuna: An empirical study of the over-representation of Target-Language Specific Features and the Unique

Items Hypotheses', in Ji, M., Oakes, M., Defeng, L. and Hareide, L. (eds.) *Corpus Methodologies Explained An empirical approach to translation studies*. London: Routledge.

Harvey, M. (2002) 'What's so special about legal translation?', *Meta*, 47(2), pp.177-185.

Hermans, T. (2007) *The Conference of the Tongues*. Manchester: St Jerome.

Hjort-Pedersen M. and Faber, D. (2010) 'Explicitation and implicitation in legal translation: A process study of trainee translators', *Meta* 2(55), pp. 237-250.

House, J. (2008) 'Beyond intervention: Universals in translation?', *Transcom*, 1(1), pp. 6-19.

Huddleston, R. (2002) 'The verb', in Huddleston, R. and Pullum, G.K. (eds.) *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press, pp. 71-212.

Inhongsakul, K., Louie, M. and Ruangjaroon, S. (2016) 'Is there a middle construction in Thai?', *The International Conference on Culture, Art, Language, and Literature in the Mekong Basin (I-CALL, 2016)*, pp. 206-221.

Iwasaki, S. (1998) 'Causative and benefactive constructions in Thai', in *Papers from the Fifth Annual Meeting of the Southeast Asian Linguistics Society*, ed. S.L. Chellian and W.J.D. Reuse, pp. 201-210. Arizona State University, Program for Southeast Asian Studies.

Iwasaki, S. and Ingkaphirom, P. (2005) *A Reference Grammar of Thai*. Cambridge: Cambridge University Press.

Kamenicka, R. (2008). 'Explicitation profile and translator style', in Pym A. and Perekrestenk A. (eds.), *Translation research projects 1*. Tarragona: Intercultural Studies Group, pp. 117-130.

Kanchanawan, N. (2006) 'Romanisation, transliteration and transcription for the globalization of the Thai language', *The Journal of the Royal Institute of Thailand*. 31(3), pp.832-841.

Kasetsiri, C. and Pongpanitanont, W. (2005) 'Robert Lingat and Thai history', in Jansang, K. and Likasitwattanakul, P (eds.) *Robert Lingat and Thai Studies*. Bangkok: Thammasart Press, pp. 55-69.

Katan, D. (2014) *Translating Cultures: An Introduction for Translators, Interpreters and Mediators*. 2<sup>nd</sup> ed. Oxon; New York: Routledge.

Kawtrakul, A., Suktarachan, M., Varasai, P. and Chanlekha, H. (2002) 'A state of the art of Thai language resources and Thai language behavior analysis and modeling', in *COLING '02 Proceedings of the 3rd workshop on Asian language resources and international standardization* 12, pp. 1-8.

Keenan, E. (1990) 'Passive in the world's languages', in Shopen, T. (ed.) *Language Typology and Syntactic Description-Clause Structure*. Volume I. Cambridge: Cambridge University Press.

Keenan, E. and Dryer, M. (2007) 'Passive in the world's languages', in Shopen, T. (ed.) *Language Typology and Syntactic Description*. Volume I. 2<sup>nd</sup> ed. Cambridge: Cambridge University Press, pp. 325-361.

Kenny, D. (2001) *Lexis and Creativity in Translation: A Corpus-Based Study*. Manchester; Northampton MA: St. Jerome.

Kenny, D. (2009) 'Corpora', in Baker, M. and Saldanha, G. (eds.) *Routledge Encyclopedia of Translation Studies*. 2<sup>nd</sup> ed. Oxford: Routledge, pp. 59-62.

Kenny, D. (2009) 'Equivalence', in Baker, M. and Saldanha, G. (eds.) *Routledge Encyclopedia of Translation Studies*. 2<sup>nd</sup> ed. London: Routledge, pp. 96-99.

Kenny, D. and Sathachai, M. (2018) 'Deontic modality in English-Thai legal translation'. Paper presented at the *Transius Conference*, 18-20 June 2018, Geneva: Université de Genève.

Kilgarrieff, A. (2018) Sketch Engine. Available at: <https://www.sketchengine.eu/>

Kilgarrieff, A., Baisa V., Bušta J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. and Suchomel, V. (2014) 'The Sketch Engine: ten years on', *Lexicography* 1 (1), pp. 7-36.

- Klaudy, K. (1996) 'Back-translation as a tool for detecting explicitation strategies in translation', in Klaudy, K., Lambert, J. and Sohar, A. (eds.) *Translation Studies in Hungary*. Budapest: Scholastica, pp. 99-114.
- Klaudy, K. (2009) 'Explicitation', in Baker, M. and Saldanha, G. (eds.) *Routledge Encyclopedia of Translation Studies*. 2<sup>nd</sup> ed. London: Routledge, pp. 104- 108.
- Knežević, B. and Brdar, I. (2011) 'Modals and modality in translation: A case study based approach', *Jezikoslovlje*, 12(2), pp. 117-145.
- Koller, W. (1979) *Einführung in die Übersetzungswissenschaft*. Heidelberg & Wiesbaden: Quelle und Meyer.
- Koller, W. (1989) 'Equivalence in Translation Theory', in Chesterman, A. (ed.) *Readings in Translation Theory*, Helsinki: Oy Finn Lectura Ab., pp. 99–104.
- Kraivixien, T. (2012) *Thai legal language*. Bangkok: Thammasart Press.
- Krogsgaard Vesterager A. (2017) 'Explicitation in legal translation — a study of Spanish-into-Danish translation of judgments', *JoSTrans. The Journal of Specialised Translation* 27(1), pp. 104-123. Available at: [http://www.jostrans.org/issue27/art\\_krogsgaard.pdf](http://www.jostrans.org/issue27/art_krogsgaard.pdf) (Accessed January 2018)
- Kruger, H. (2012) 'A corpus-based study of the mediation effect in translated and edited language', *Target*, 24(2), pp. 355-388.
- Kullavanijava, P. (1974) *Transitive Verb in Thai*. PhD thesis. University of Hawaii.
- Kurzon, D. 1986. *It Is Hereby Performed... Explorations in Legal Speech Acts*. Amsterdam/Philadelphia: John Benjamins.
- Laviosa, S. (1998) 'Core patterns of lexical use in a comparable corpus of English narrative prose', *Meta*, 43(4), pp. 557-550.
- Laviosa, S. (2002) *Corpus-Based Translation Studies: Theory, Findings, Applications*. Amsterdam: Rodopi.
- Lembersky, G. (2013) *The Effect of Translationese on the Statistical Machine Translation*. PhD thesis. University of Haifa.

Lembersky, G., Ordan, N. and Wintner, S. (2012) 'Language models for machine translation: Original vs. translated texts', *Computational Linguistics* 38(4), pp.799-825.

Lembersky, G., Ordan, N. and Wintner, S. (2013) 'Improving statistical machine translation by adapting translation models to translationese', *Computational Linguistics* 39(4), pp. 999-1023.

Loock, R. (2016) *La Traductologie de Corpus*. Villeneuve d'Ascq: Presses Universitaires du Septentrion.

Loos, T. (1999) *Subject Siam: Family, Law, and Colonial Modernity in Thailand*. Ithaca; London: Cornell University Press.

Low, J. (1828) *A Grammar of Thai or Siamese Language*. Calcutta: Baptist

Marinetti, C. (2008) 'Translating law by Cao, D. (2007): A book review', *The International Journal of Speech, Language and the Law*, 15(1), pp. 121-125.

Martikainen, H. (2016) 'A functional approach to translation quality assessment: Categorizing sources of translation bias in medical abstracts', *Linguistica Antverpiensia*, 14 (New Series: Themes in Translation Studies), pp. 153-174.

Mattila, H. (2006) *Comparative Legal Linguistics*. Aldershot: Ashgate.

Mauranen, A. (2007) 'Universal tendencies in translation', in Anderman, G.M. and Margaret, R. (eds.) *Incorporating Corpora. The Linguist and the Translator*. Clevedon: Multilingual Matters, pp. 32-48.

Mellinkoff, D. (2004) *The Language of the Law*. Oregon: Resource Publications.

Montgomery, S. (2000) *Science in Translation: Movements of Knowledge through Cultures and Time*. London; Chicago: University of Chicago Press.

Nadrag, L., Buzarna-Tihenea, A. and Stan, A. (2013) 'A quest for a model: The globalization of English legal language', *Contemporary Readings in Law & Social Justice*, 5(2), pp. 844-853.

Narrog, H. (2009) *Modality in Japanese*. Amsterdam; Philadelphia: John Benjamins.



- Narrog, H. (2016) 'The expression of non-epistemic modal categories' in Nuyts, J. and Van Der Auwera, J. (eds.) *The Oxford handbook of modality and mood*. 1<sup>st</sup> ed. Oxford: Oxford University Press, pp. 89-116.
- Neubert, A. (1997) 'Postulates for a theory of translation', in Danks, J. et al (eds.) *Cognitive Processes in Translating and Interpreting*. Thousand Oaks, CA: Sage, pp. 1-24.
- Neubert, A. and Shreve, G. (1992) *Translating as Text*. Kent; Ohio: The Kent State University Press.
- Newman, A. (1994) 'Translation Equivalence: Nature', in Asher, R. E. and Simpson, J. M. Y. (eds.) *The Encyclopedia of Language and Linguistics*. Oxford & New York: Pergamon Press, pp. 469-700.
- Nida, E. (1964) *Toward a Science of Translating*. Leiden: E. J. Brill.
- Olohan, M. (2004) *Introducing Corpora in Translation Studies*. London; New York: Routledge.
- Olohan, M. and Baker, M. (2000) 'Reporting *that* in translated English: Evidence for subconscious processes of explicitation?', *Across Languages and Cultures*, 1, pp. 141-158.
- Øverås, L. (1998) 'In search of the third code: An investigation of norms in literary translation', *Meta*, 43, pp. 571-588.
- Palmer, D. (2010) 'Text preprocessing', in Indurkha, N. and Damerau, F. J. (eds.) *Handbook of Natural Language Processing*, 2<sup>nd</sup> ed, Boca Raton FLA: CRC Press, pp. 9-30.
- Palmer, F.R. (1986) *Mood and Modality*. Cambridge: Cambridge University Press.
- Pápai, V. (2004) 'Explicitation. A universal of translated text?', in Mauranen, A. and Kujamäki, P. (eds) *Translation Universals: Do They Exist?* Amsterdam; Philadelphia: John Benjamins, pp. 143-164.

- Patpong, P. (2006) *A Systemic Functional Interpretation of Thai Grammar: An Exploration of Thai Narrative Discourse*. PhD thesis. Macquarie University.
- Paunio, E. and Lindroos-Hovinheimo, S. (2010) 'Taking language seriously: An analysis of linguistic reasoning and its implications in EU law', *European Law Journal*, 16(4), pp. 395-416.
- Pigeon, L. (1978) *Rédaction et interprétation des lois*. Québec: Éditeur officiel du Québec.
- Pommer, S.E. (2008) 'No creativity in legal translation?', *Babel*, 54(4), pp. 355-368.
- Powell, R. and Hashim, A. (2011) 'Language disadvantage in Malaysian litigation and arbitration', *World Englishes*, 30(1), pp. 92-105.
- Prasithrathsint, A. (1988) 'Change in the passive constructions in Standard Thai from 1802 to 1982', *Language Sciences*, 10(2), pp. 363-393.
- Prasithrathsint, A. (2003) 'A typological approach to the passive in Thai', *Manusya: Journal of Humanities*, (6), pp. 1-17.
- Prasithrathsint, A. (2006) 'Development of the thùuk passive marker in Thai', in Abraham, W. and Leisiö, L. (eds.) *Passivization and Typology: Form and Function*. Amsterdam; Philadelphia: John Benjamins, pp. 115-131.
- Prasithrathsint, A. (2009) Cohesion. In *Thai Language Standard: Type of Words, Phrases, Sentences and Cohesion*. Bangkok: Business Organisation of the Office of the Welfare Promotion Commission for Teachers and Educational Personnel.
- Pym, A. (2005) 'Explaining explicitation' in Klaudy, K., Károly, K. and Fóris, Á. (eds.) *New Trends in Translation Studies: In Honour of Kinga Klaudy*. Budapest: Akadémia Kiadó, pp. 29-34.
- Pym, A. (2015) 'Translating as risk management', *Journal of Pragmatics*, 85, pp. 67-80.
- Rachabanditsathan Dictionary* (1999) Bangkok: Nanmee Book Publishing.

Sager, J. (1993) *Language, Engineering and Translation: Consequences of Automation*. Philadelphia: Benjamins.

Saldanha, G. (2008) 'Explicitation Revisited: Bringing the Reader into the Picture', *trans-kom* 1 (1), pp 20-35. Available at: [http://www.trans-kom.eu/bd01nr01/trans-kom\\_01\\_01\\_03\\_Saldanha\\_Explicitation.20080707.pdf](http://www.trans-kom.eu/bd01nr01/trans-kom_01_01_03_Saldanha_Explicitation.20080707.pdf). (Accessed 4 May 2015)

Salkie, R. (2010) 'Will: Tense or modal or both?', *English Language and Linguistics*, 14(2), pp. 187-215.

Šarčević, S. (1997) *New Approach to Legal Translation*. The Hague: Kluwer Law International.

Šarčević, S. (2012) 'Challenges to the legal translator', in Tiersma, P. and Solan, L. (eds.) *The Oxford Handbook of Language and Law*. Oxford: Oxford University Press, pp.187-199.

Schane, S. (2008) *Language and the law*. London: Continuum .

Singhapreecha, P. (2007). 'Book review: A reference grammar of Thai', *Lingua* 117(8), pp. 1497-1512.

Smalley, W. (1994) *Linguistic Diversity and National Unity: Language Ecology in Thailand*. Chicago; London: The University of Chicago Press.

Smith, R. and Frawley, W. (2009) 'Conjunctive cohesion in four English genres', *Text - Interdisciplinary Journal for the Study of Discourse*, 3(4), pp. 347-374.

Solan, L.M. (1993) *Language of Judges*. Chicago: University of Chicago Press.

Sornlertlamvanich, V., Charoenporn, T. and Isahara, H. (1997) 'ORCHID: Thai part-of-speech tagged corpus', National Electronics and Computer Technology Centre Technical Report, pp. 5-19.

Srioutai, J. (2004) 'The Thai *Cla*: A marker of tense or modality?', *Camling* 2, pp. 100-107.

- Sriyapai, W. (2013) *Thai Linguistics*. Nonthaburi: Sampachanya Publishing.
- Sternberg, R. (ed.) (1999) *Handbook of Creativity*. Cambridge: Cambridge University Press.
- Stubbs, M. (1996) *Text and Corpus Analysis*. Oxford; Cambridge MA: Basil Blackwell.
- Sudmuk, C. (2003) 'The thuk construction in Thai', in Butt, M. and King, T. H. (eds.) *The LFG03 Conference University at Albany, State University of New York*. Stanford, United States: CSLI Publication, pp.402-423.
- Sumner, C. (1979) *Reading Ideologies*. London; New York; San Francisco: Academic Press.
- Teich, E. (2003) *Cross-linguistic variation in system and text. A method for the investigation of translations and comparable texts*. Berlin: Mouton de Gruyter.
- Thomas, J. (2016) *Discovering English with Sketch Engine*. 2<sup>nd</sup> ed. Brno: Versatile.
- Tiersma, P. (1999) *Legal Language*. Chicago: The University of Chicago Press.
- Tirkkonen-Condit, S. (2002) 'Translationese – a myth or an empirical fact?', *Target* 14(2), pp. 207-220.
- Tirkkonen-Condit, S. (2004) 'Unique items – over- or under-represented in translated language? in Mauranen, A. and Kuusimäki, P. (eds.) *Translation Universals: Do They Exist?* Amsterdam; Philadelphia: John Benjamins, pp. 177-184.
- Toury, G. (1995) *Descriptive Translation Studies and Beyond*. Amsterdam: John Benjamins.
- Uwanno, B. (2000) *Administrative Law*. Bangkok: Netibhandisapa.
- Vienna Convention on the Law of treaties* (1969).
- Vilinsky, B. Á. M. (2012) 'On the lower frequency of occurrence of Spanish verbal periphrases in translated texts as evidence for the unique items hypothesis', *Across Languages and Cultures*, 13(2), pp. 197–210.

Vinay, J. and Darbelnet, J. (1995) *Comparative Stylistics of French and English: A Methodology for Translation*, translated and edited by Sager J. C. and Hamel M. J. Amsterdam: John Benjamins.

von Fintel, K. (2006) 'Modality and Language', in Borchert, D. M. (ed) *Encyclopedia of Philosophy*, 2<sup>nd</sup> ed, Volume 10. Detroit: MacMillan Reference USA, pp. 20–27

Wakabayashi, J. (2014) 'Positionality, representation and questions on translation in Thailand', the *First International Conference on Translation Studies: Translating Asia: Then and Now*, 5-6 June 2014, Bangkok, Thailand. Available at: <http://icts.utcc.ac.th/wp-content/uploads/2014/12/Positionality-Representation-and-Questions-on-Translation-in-Thailand.pdf> (Accessed 29 September 2017)

Weisflog, W. E. (1987) 'Problems of legal translation', *Swiss Reports Presented at the XIIth International Congress of Comparative Law*. pp.179-218.

Weston, M. (1991) *An English Reader's Guide to the French Legal System*. New York; Oxford : Berg.

Wilcox, S. and Shaffer, B. (2008) 'Modality in American sign language', in Frawley, W., Eschenroede, E., Mills, S. and Nguyen, T. (eds.) *The Expression of Modality*. Berlin: Walter De Gruyter, pp. 207-234.

Williams, C. (2009) 'Legal English and the modal revolution', in Salkie, R., Busuttil, P. and Van Der Auwera, J. (eds.) *Modality in English: Theory and Description*. Berlin: Mouton De Gruyter, pp. 199-210.

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), pp.5-35.



## Appendices

## Appendix A: Lists of Texts in the Parallel Corpus

(Note: the total word counts reported here differ from those reported in the body of the thesis as this word count is provided by Microsoft Word based on Microsoft's proprietary word tokenizers for English and Thai, whereas the word counts reported in the body of the thesis are based on the tokenizers used by Sketch Engine.)

	English Name	Year	No. of words	Thai Name	No. of words
1.	Charter of The ASEAN	2007	5,036	กฎบัตรอาเซียน	6,090
2.	ASEAN Human Rights Declaration	2012	2,423	ปฏิญญาอาเซียนว่าด้วยสิทธิมนุษยชน	2,908
3.	International Convention on Civil and Political Rights – ICCPR)	1976	6,849	กติการะหว่างประเทศว่าด้วยสิทธิพลเมืองและสิทธิทางการเมือง	7,468
4.	International Convention on Economic, Social and Cultural Rights – ICESCR)	1976	3,573	กติการะหว่างประเทศว่าด้วยสิทธิทางเศรษฐกิจ สังคม และวัฒนธรรม	3,831
5.	Convention of the Elimination of All Forms of Discrimination against Women	1979	4,328	อนุสัญญาว่าด้วยการจัดการเลือกปฏิบัติต่อสตรีในทุกรูปแบบและพิธีสารเลือกรับของอนุสัญญาว่าด้วยการจัดการเลือกปฏิบัติต่อสตรีในทุกรูปแบบ	5,243
6.	Convention on the Rights of the Child	1989	7,410	อนุสัญญาว่าด้วยสิทธิเด็ก	8,540
7.	Convention on the Elimination of all Forms of Racial Discrimination – CERD	1969	4,643	อนุสัญญาระหว่างประเทศว่าด้วยการจัดการเลือกปฏิบัติทางเชื้อชาติในทุกรูปแบบ	5,247
8.	Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or	1990	5,230	อนุสัญญาต่อต้านการทรมานและการประติบัติหรือการลงโทษที่โหดร้ายไม่มนุษยธรรมหรือที่ข้ำยักยัก	6,395



	Punishment				
9.	Convention on the Rights of Person with Disabilities	2006	9,565	อนุสัญญาว่าด้วยสิทธิคนพิการ	11,505
10.	International Convention for the protection of All Persons from Enforced Disappearance	2006	6,941	อนุสัญญาระหว่างประเทศ ว่าด้วยการคุ้มครองบุคคลทุกคนจากการหายสาบสูญโดยถูกบังคับ	7,594
11.	Universal Declaration of Human Rights	1948	1,759	ปฏิญญาสากลว่าด้วยสิทธิมนุษยชน	2,106
12.	Convention on International Trade in Endangered Species of Wild Fauna and Flora	1975	6,692	อนุสัญญาไซเตส	7,694
13.	United Nations Framework Convention on Climate Change	1992	8,114	อนุสัญญาสหประชาชาติว่าด้วยการเปลี่ยนแปลงสภาพ	9,942
14.	Kyoto Protocol to United Nations Framework Convention on Climate Change	1997	8,278	พิธีสารเกียวโต	9,671
15.	Convention on Biological Diversity	1992	7,859	อนุสัญญาว่าด้วยความหลากหลายทางชีวภาพ	9,640
16.	Double Tax Agreements with Thailand (45 countries)	(1992-2010)	11,032	อนุสัญญาภาษีซ้อน	12,970
17.	Basel Convention on the Control of Transboundary Movements of Hazardous Wastes and Their Disposal	1989	8,496	อนุสัญญาบาเซล	10,035
18.	United Nations Convention to Combat Desertification in Countries Experiencing	1994	10,052	อนุสัญญาสหประชาชาติว่าด้วยการต่อต้านการแปรสภาพเป็นทะเลทราย ในประเทศที่ประสบภัยแล้งอย่างรุนแรงและ/หรือการแปรสภาพเป็นทะเลทราย	12,238

	Serious Drought and/or Desertification, Particularly in Africa			โดยเฉพาะในทวีปแอฟริกา	
19.	Rotterdam Convention	1998	6,654	อนุสัญญารอตเตอร์ดัม	7,335
20.	Stockholm Convention on Persistent Organic	2001	8,982	อนุสัญญาสตอกโฮล์ม	10,910
21.	Minamata Convention on Mercury	2013	10,818	อนุสัญญามินามาตะว่าด้วยปรอท	12,890
22.	Montreal Convention (Aviation)	1999	8,171	อนุสัญญามอนทรีออล	9,794
23.	International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families	1990	14,084	อนุสัญญาว่าด้วยการคุ้มครองสิทธิของแรงงานอพยพและสมาชิกครอบครัว	15,255
24.	Convention Concerning the Protection of the World Cultural and Natural Heritage	1972	4,652	อนุสัญญาว่าด้วยการคุ้มครองมรดกโลกทางวัฒนธรรมและทางธรรมชาติ	5,546
25.	Convention on Protection of Children and Co-operation in respect of Intercountry Adoption	1993	4,202	อนุสัญญาว่าด้วยการคุ้มครองเด็กและความร่วมมือเกี่ยวกับการรับบุตรบุญธรรมระหว่างประเทศ	4,490
	Total		175,816		205,337

## Appendix B: List of Thai Legislative Texts for Thai Monolingual Corpus

(Note: the total word count reported here differs from that reported in the body of the thesis as this word count is provided by Microsoft Word based on Microsoft's proprietary word tokenizer for Thai, whereas the word count reported in the body of the thesis (1,173,485 words) is based on the Swath tokenizer used by Sketch Engine.)

	Name	Year	No. of words
1.	รัฐธรรมนูญแห่งราชอาณาจักรไทยฉบับชั่วคราว (พุทธศักราช ๒๕๕๗) ๑[	2014	9,679
2.	พระราชบัญญัติให้ใช้ประมวลกฎหมายที่ดิน พ.ศ. ๒๔๘๗	1954	19,371
3.	พระราชบัญญัติให้ใช้ประมวลกฎหมายวิธีพิจารณาความอาญา พุทธศักราช ๒๔๗๗	1934 (amended in 2015)	42,498
4.	พระราชบัญญัติให้ใช้ประมวลกฎหมายวิธีพิจารณาความแพ่งพุทธศักราช ๒๔๗๗	1934 (amended in 2015)	76,821
5.	พระราชบัญญัติให้ใช้ประมวลกฎหมายอาญา พ.ศ. ๒๔๘๕	1956 (amended in 2016)	44,350
6.	ประมวลกฎหมายอาญาทหาร	1911 (amended in 1964)	5,964
7.	ประมวลกฎหมายแพ่งและพาณิชย์	1925 (amended in 2015)	133,616
8.	ประมวลรัษฎากร	1938 (amended in 2016)	73,858
9.	พระราชกำหนดการนำคนต่างด้าวมาทำงานกับนายจ้างในประเทศ พ.ศ. ๒๕๕๕	2016	6,084
10.	พระราชบัญญัติการให้บริการด้านนิติวิทยาศาสตร์ พ.ศ. ๒๕๕๕	2016	1,467
11.	พระราชบัญญัติการป้องกันและแก้ไขปัญหาดังกล่าวในวัยรุ่น พ.ศ. ๒๕๕๕	2016	2,869
12.	พระราชบัญญัติควบคุมการขอยาน พ.ศ. ๒๕๕๕	2016	2,539

13.	พระราชบัญญัติคุ้มครองการดำเนินงานของสำนักเลขานุการองค์กรสำรองข้าว ฉุกเฉินของอาเซียนบวกสามพ.ศ .๒๕๕๕	2016	660
14.	พระราชบัญญัติคุ้มครองการดำเนินงานของทบวงการพลังงานหมุนเวียนระหว่าง ประเทศพ.ศ .๒๕๕๕	2016	492
15.	พระราชบัญญัติจัดตั้งศาลอาญาคดีทุจริตและประพฤติมิชอบพ.ศ .๒๕๕๕	2016	2,466
16.	พระราชบัญญัติพลังงานนิวเคลียร์เพื่อสันติพ.ศ .๒๕๕๕	2016	15,971
17.	พระราชบัญญัติสถานประกอบการเพื่อสุขภาพ พ.ศ .๒๕๕๕	2016	4,821
18.	พระราชบัญญัติให้อำนาจปฏิบัติการเกี่ยวกับธนาคารเพื่อการลงทุนในโครงสร้าง พื้นฐานเอเชีย พ.ศ .๒๕๕๕	2016	760
19.	พระราชบัญญัติมหาวิทยาลัยศรีนครินทรวิโรฒ พ.ศ .๒๕๕๕	2016	9,397
20.	พระราชบัญญัติว่าด้วยการออกเสียงประชามติร่างรัฐธรรมนูญ พ.ศ .๒๕๕๕	2016	8,728
21.	พระราชบัญญัติวิธีพิจารณาคดีค้ามนุษย์ พ.ศ .๒๕๕๕	2016	4,198
22.	พระราชบัญญัติส่งเสริมและรักษามรดกภูมิปัญญาทางวัฒนธรรม พ.ศ .๒๕๕๕	2016	2,789
23.	พระราชกำหนดการบินพลเรือนแห่งประเทศไทยพ.ศ .๒๕๕๘	2015	5,854
24.	พระราชกำหนดการประมง พ.ศ .๒๕๕๘	2015	20,391
25.	พระราชบัญญัติกองทุนพัฒนาระบบสถาบันการเงินเฉพาะกิจ พ.ศ .๒๕๕๘	2015	1,958
26.	พระราชบัญญัติกองทุนพัฒนาสื่อปลอดภัยและสร้างสรรค์ พ.ศ .๒๕๕๘	2015	4,020
27.	พระราชบัญญัติกองทุนยุติธรรม พ.ศ .๒๕๕๘	2015	2,976
28.	พระราชบัญญัติการกีฬาแห่งประเทศไทย พ.ศ .๒๕๕๘	2015	12,070
29.	พระราชบัญญัติการบริหารทุนหมุนเวียน พ.ศ .๒๕๕๘	2015	3,350
30.	พระราชบัญญัติการปฏิบัติหน้าที่ของคณะกรรมการที่แต่งตั้งตามประกาศและ คำสั่งของคณะรักษาความสงบแห่งชาติฉบับที่ พ.ศ .๒๕๕๘	2015	1,452
31.	พระราชบัญญัติการยางแห่งประเทศไทย พ.ศ .๒๕๕๘	2015	6,681
32.	พระราชบัญญัติการรับขนทางอากาศระหว่างประเทศ พ.ศ .๒๕๕๘	2015	7,188
33.	พระราชบัญญัติการอำนวยความสะดวกในการพิจารณาอนุญาตของทางราชการ พ.ศ .๒๕๕๘	2015	2,874
34.	พระราชบัญญัติกำลังพลสำรอง พ.ศ .๒๕๕๘	2015	3,456
35.	พระราชบัญญัติกลังสินค้า ไซโล และห้องเย็น พ.ศ .๒๕๕๘	2015	6,573

36.	พระราชบัญญัติควบคุมคุณภาพอาหารสัตว์ พ.ศ .๒๕๕๘	2015	9,232
37.	พระราชบัญญัติความเท่าเทียมระหว่างเพศ พ.ศ .๒๕๕๘	2015	3,639
38.	พระราชบัญญัติคุ้มครองเด็กที่เกิดโดยอาศัยเทคโนโลยีช่วยการเจริญพันธุ์ทางการแพทย์ พ.ศ .๒๕๕๘	2015	4,508
39.	พระราชบัญญัติเครื่องสำอาง พ.ศ .๒๕๕๘	2015	8,716
40.	พระราชบัญญัติงาช้าง พ.ศ .๒๕๕๘	2015	1,923
41.	พระราชบัญญัติจตุรพักดินเพื่อกษัตริย์ พ.ศ .๒๕๕๘	2015	9,338
42.	พระราชบัญญัติธุรกิจรักษาความปลอดภัย พ.ศ .๒๕๕๘	2015	6,028
43.	พระราชบัญญัติภาษีการรับมรดก พ.ศ .๒๕๕๘	2015	4,219
44.	พระราชกำหนดยกเว้นและสนับสนุนการปฏิบัติการเกี่ยวกับภาษีอากรตามประมวลรัษฎากร พ.ศ .๒๕๕๘]๑[	2015	1,198
45.	พระราชบัญญัติราชบัณฑิตยสภา พ.ศ .๒๕๕๘	2015	3,418
46.	พระราชบัญญัติแรงงานทางทะเล พ.ศ .๒๕๕๘	2015	12,217
47.	พระราชบัญญัติโรคระบาดสัตว์ พ.ศ .๒๕๕๘	2015	8,516
48.	พระราชบัญญัติสัตว์เพื่อนทางวิทยาศาสตร์ พ.ศ .๒๕๕๘	2015	5,631
49.	พระราชบัญญัติส่งเสริมการบริหารจัดการทรัพยากรทางทะเลและชายฝั่ง พ.ศ . ๒๕๕๘	2015	4,376
50.	พระราชบัญญัติหลักประกันทางธุรกิจ พ.ศ .๒๕๕๘	2015	10,283
51.	พระราชบัญญัติหอพัก พ.ศ .๒๕๕๘	2015	8,023
52.	พระราชบัญญัติการคุ้มครองคนไร้ที่พึ่ง พ.ศ .๒๕๕๗	2014	3,248
53.	พระราชบัญญัติการรับคนโดยสารทางถนนระหว่างประเทศ พ.ศ .๒๕๕๗	2014	3,134
54.	พระราชบัญญัติป้องกันการทารุณกรรมและการจัดสวัสดิภาพสัตว์ พ.ศ .๒๕๕๗	2014	3,280
55.	พระราชบัญญัติกองทุนเพื่อผู้เคยเป็นสมาชิกรัฐสภา พ.ศ .๒๕๕๖	2013	2,207
56.	พระราชบัญญัติการอำนวยความสะดวกในการขนส่งข้ามพรมแดน พ.ศ .๒๕๕๖	2013	1,511
57.	พระราชบัญญัติคุ้มครองการประทุษร้ายอันเนื่องด้วยการค้าระหว่างประเทศ ซึ่งชนิดสัตว์ป่าและพืชป่าที่ใกล้สูญพันธุ์ ครั้งที่ ๑๖ ในประเทศไทย พ.ศ . ๒๕๕๖	2013	853
58.	พระราชบัญญัติป้องกันและปราบปรามการมีส่วนร่วมในองค์กรอาชญากรรมข้าม	2013	4,504

	ชาติ พ.ศ .๒๕๕๖		
59.	พระราชบัญญัติป้องกันและปราบปรามการสนับสนุนทางการเงินแก่การก่อการร้าย พ.ศ .๒๕๕๖	2013	2,950
60.	พระราชบัญญัติวิชาชีพการสาธารณสุขชุมชน พ.ศ .๒๕๕๖	2013	5,895
61.	พระราชบัญญัติวิชาชีพการแพทย์แผนไทย พ.ศ .๒๕๕๖	2013	7,895
62.	พระราชบัญญัติวิชาชีพสังคมสงเคราะห์ พ.ศ .๒๕๕๖	2013	5,346
63.	พระราชบัญญัติส่งเสริมกีฬาอาชีพ พ.ศ .๒๕๕๖	2013	6,625
64.	พระราชบัญญัติการรับขนของทางถนนระหว่างประเทศ พ.ศ .๒๕๕๖	2013	5,309
65.	พระราชบัญญัติการให้เอกชนร่วมลงทุนในกิจการของรัฐ พ.ศ .๒๕๕๖	2013	8,269
66.	พระราชกำหนดกองทุนส่งเสริมการประกันภัยพิบัติ พ.ศ .๒๕๕๕	2012	1,801
67.	พระราชกำหนดปรับปรุงการบริหารหนี้เงินกู้ที่กระทรวงการคลังกู้เพื่อช่วยเหลือกองทุนเพื่อการฟื้นฟูและพัฒนาระบบสถาบันการเงิน พ.ศ .๒๕๕๕	2012	1,881
68.	พระราชบัญญัติสถาบันดนตรีกัลยาณีวัฒนา พ.ศ .๒๕๕๕	2012	6,680
69.	พระราชกำหนดให้อำนาจกระทรวงการคลังกู้เงินเพื่อการวางระบบบริหารจัดการน้ำและสร้างอนาคตประเทศพ.ศ .๒๕๕๕	2012	1,171
70.	พระราชบัญญัติความร่วมมือระหว่างประเทศในทางแพ่งเกี่ยวกับการละเมิดสิทธิควบคุมดูแลเด็ก พ.ศ .๒๕๕๕	2012	2,754
71.	พระราชบัญญัติกองทุนการออมแห่งชาติ พ.ศ .๒๕๕๔	2011	6,023
72.	พระราชบัญญัติกองทุนสงเคราะห์เกษตรกร พ.ศ .๒๕๕๔	2011	2,891
73.	พระราชบัญญัติคณะกรรมการปฏิรูปกฎหมาย พ.ศ .๒๕๕๓	2010	3,355
74.	พระราชบัญญัติระเบียบข้าราชการฝ่ายอัยการ พ.ศ .๒๕๕๓	2010	12,622
75.	พระราชบัญญัติวัฒนธรรมแห่งชาติ พ.ศ .๒๕๕๓	2010	2,455
76.	พระราชบัญญัติศาลเยาวชนและครอบครัวและวิธีพิจารณาคดีเยาวชนและครอบครัว พ.ศ .๒๕๕๓	2010	25,993
77.	พระราชบัญญัติองค์กรจัดสรรคลื่นความถี่และกำกับการประกอบกิจการวิทยุกระจายเสียง วิทยุโทรทัศน์ และกิจการโทรคมนาคม พ.ศ .๒๕๕๓	2010	15,809
78.	พระราชบัญญัติองค์กรอัยการและพนักงานอัยการ พ.ศ .๒๕๕๓	2010	3,563
79.	พระราชบัญญัติการบริหารราชการจังหวัดชายแดนภาคใต้ พ.ศ .๒๕๕๓	2010	4,877

80.	พระราชบัญญัติประกอบรัฐธรรมนูญว่าด้วยการออกเสียงประชามติ พ.ศ . ๒๕๕๒	2009	6,813
81.	พระราชบัญญัติประกอบรัฐธรรมนูญว่าด้วยผู้ตรวจการแผ่นดิน พ.ศ . ๒๕๕๒	2009	5,024
82.	พระราชกำหนดให้อำนาจกระทรวงการคลังกู้เงินเพื่อฟื้นฟูและเสริมสร้างความมั่นคงทางเศรษฐกิจ พ.ศ . ๒๕๕๒	2009	1,094
83.	พระราชบัญญัติการจัดการศึกษาสำหรับคนพิการ พ.ศ . ๒๕๕๑	2008	3,798
84.	พระราชบัญญัติการดูแลผลประโยชน์ของคู่สัญญา พ.ศ . ๒๕๕๑	2008	4,549
85.	พระราชบัญญัติการทำงานของคนต่างด้าว พ.ศ . ๒๕๕๑	2008	5,987
86.	พระราชบัญญัติการประกอบกิจการกระจายเสียงและกิจการโทรทัศน์ พ.ศ . ๒๕๕๑	2008	9,675
87.	พระราชบัญญัติการมาตรฐานแห่งชาติ พ.ศ . ๒๕๕๑	2008	5,593
88.	พระราชบัญญัติการรักษาความมั่นคงภายในราชอาณาจักร พ.ศ . ๒๕๕๑	2008	3,602
89.	พระราชบัญญัติการแพทย์ฉุกเฉิน พ.ศ . ๒๕๕๑	2008	4,874
90.	พระราชบัญญัติการอาชีวศึกษา พ.ศ . ๒๕๕๑	2008	6,523
91.	พระราชบัญญัติคณะกรรมการอาหารแห่งชาติ พ.ศ . ๒๕๕๑	2008	2,307
92.	พระราชบัญญัติควบคุมเครื่องดัดแปลงฮาร์ดแวร์ พ.ศ . ๒๕๕๑	2008	4,396
93.	พระราชบัญญัติความรับผิดชอบต่อความเสียหายที่เกิดขึ้นจากสินค้าที่ไม่ปลอดภัย พ.ศ . ๒๕๕๑	2008	1,784
94.	พระราชบัญญัติคุ้มครองซากดึกดำบรรพ์ พ.ศ . ๒๕๕๑	2008	6,880
95.	พระราชบัญญัติเครื่องมือแพทย์ พ.ศ . ๒๕๕๑	2008	13,328
96.	พระราชบัญญัติโคนมและผลิตภัณฑ์นม พ.ศ . ๒๕๕๑	2008	1,545
97.	พระราชบัญญัติจัดระเบียบราชการกระทรวงกลาโหม พ.ศ . ๒๕๕๑	2008	4,511
98.	พระราชบัญญัติธุรกิจนำเข้าและส่งออก พ.ศ . ๒๕๕๑	2008	10,318
99.	พระราชบัญญัตินโยบายการท่องเที่ยวแห่งชาติ พ.ศ . ๒๕๕๑	2008	3,323
100.	พระราชบัญญัติป้องกันและปราบปรามการค้ามนุษย์ พ.ศ . ๒๕๕๑	2008	8,192
101.	พระราชบัญญัติพัฒนาที่ดิน พ.ศ . ๒๕๕๑	2008	2,856
102.	พระราชบัญญัติภาพยนตร์และวีดิทัศน์ พ.ศ . ๒๕๕๑	2008	7,543
103.	พระราชบัญญัติมาตรการของฝ่ายบริหารในการป้องกันและปราบปรามการทุจริต	2008	10,064

	พ.ศ .๒๕๕๑		
104.	พระราชบัญญัติมาตรฐานสินค้าเกษตร พ.ศ .๒๕๕๑	2008	7,426
105.	พระราชบัญญัติระเบียบข้าราชการพลเรือน พ.ศ .๒๕๕๑	2008	17,869
106.	พระราชบัญญัติโรงเรียนนายร้อยตำรวจ พ.ศ .๒๕๕๑	2008	3,223
107.	พระราชบัญญัติลูกเสือ พ.ศ .๒๕๕๑	2008	6,165
108.	พระราชบัญญัติวิธีพิจารณาคดีผู้บริโภค พ.ศ .๒๕๕๑	2008	7,557
109.	พระราชบัญญัติสถาบันคุ้มครองเงินฝาก พ.ศ .๒๕๕๑	2008	5,992
110.	พระราชบัญญัติสภาพัฒนาการเมือง พ.ศ .๒๕๕๑	2008	4,292
111.	พระราชบัญญัติส่งผู้ร้ายข้ามแดน พ.ศ .๒๕๕๑	2008	4,288
112.	พระราชบัญญัติส่งเสริมการศึกษานอกระบบและการศึกษาตามอัธยาศัย พ.ศ . ๒๕๕๑	2008	3,047
113.	พระราชบัญญัติส่งเสริมวิชาชีพวิทยาศาสตร์และเทคโนโลยี พ.ศ .๒๕๕๑	2008	5,944
114.	พระราชบัญญัติส่งเสริมศิลปกรรมร่วมสมัย พ.ศ .๒๕๕๑	2008	2,439
115.	พระราชบัญญัติองค์การกระจายเสียงและแพร่ภาพสาธารณะแห่งประเทศไทย พ.ศ .๒๕๕๑	2008	7,245
116.	พระราชบัญญัติคณะกรรมการกำกับและส่งเสริมการประกอบธุรกิจประกันภัย พ.ศ .๒๕๕๐	2008	4,633
117.	พระราชบัญญัติการทางพิเศษแห่งประเทศไทย พ.ศ .๒๕๕๐	2007	5,688
118.	พระราชบัญญัติการประกอบกิจการพลังงาน พ.ศ .๒๕๕๐	2007	17,060
119.	พระราชบัญญัติคุ้มครองการดำเนินงานขององค์การความร่วมมือด้านอากาศแห่ง เอเชียแปซิฟิก พ.ศ .๒๕๕๐	2007	5,672
120.	พระราชบัญญัติคุ้มครองการดำเนินงานขององค์การดาวเทียมเคลื่อนที่ระหว่าง ประเทศ พ.ศ .๒๕๕๐	2007	4,230
121.	พระราชบัญญัติคุ้มครองผู้ถูกระทำด้วยความรุนแรงในครอบครัว พ.ศ .๒๕๕๐	2007	2,772
122.	พระราชบัญญัติจดทะเบียนการพิมพ์ พ.ศ .๒๕๕๐	2007	2,551
123.	พระราชกำหนดการบริหารราชการในสถานการณ์ฉุกเฉิน พ.ศ .๒๕๔๘	2005	3,138
	Total		1,058,217



## Appendix C: The third Royal Institute Transcription System

### *Thai Consonants*

Thai consonant	Initial	Final
ก	K-	-K
ข ฃ ฅ	KH-	-K
ง	NG-	-NG
จ	CH-	-T
ฉ ช ฌ	CH-	-T
ญ	Y-	-N
ด ฎ ฐ	D-	-T
ต ฏ	T-	-T
ถ ฐ ท ฑ ฒ ฒ	TH-	-T
น ณ	N-	-N
บ	B-	-P
ป	P-	-P
ผ พ ภ	PH-	-P
ฝ ฟ	F-	-P
ม	M-	-M
ย	Y-	-
ร	R-	-N
ล ฬ	L-	-N
ว	W-	-
ซ ฌ ฌ ฌ	S-	-T
ห ฮ	H-	-
อ	-	-

### Thai Vowels

Vowels	Transliteration
ะ ą- ั - -	a
ำ-	am
ิ- ี -	i
อ- ัว- ือ-	ue
ุ- ู, -	u
เ- ะ ะ -	e
แ- ะ แ-	ae
โ- ะ- ะ- ะ- ะ- ะ-	o
เ- ะ เ- ะ-	oe
เ- ะ- ะ- ะ-	ia
เ- ะ- ะ- ะ-	uea
ว- ัว- ัว- ะ-	ua
เ- ะ- ะ- ะ- ะ- ะ-	ai
เ- ะ- ะ-	ao
เ- ะ- ะ-	ui
เ- ะ- ะ-	oi
เ- ะ-	iu
เ- ะ- ะ-	eo
เ- ะ-	oei
เ- ะ- ะ-	ueai
เ- ะ-	uai
เ- ะ-	aeo
เ- ะ- ะ-	ieo

## **Appendix D: Experience of Thai Corpus Compilation**

For the benefit of future researchers, we provide a brief summary here of our experience of Thai corpus compilation with the main focus on the most significant problems that we encountered while building and processing the Thai corpora and how we overcame them. We hope that other researchers will be able to extrapolate recommended practices from this summary.

### **Corpus Building**

For the English-Thai parallel corpus compilation, as mentioned earlier, after obtaining texts from the web sources, we converted the files to an appropriate format, in this case, Text Only format. In fact, Sketch Engine supports multiple formats (such as .doc, .docx, .htm, .html, and .pdf). However, it is much easier to align English texts with Thai texts if files are converted into Text Only format.

For the English original texts, there were very few conversion errors to be corrected. However, for the Thai translated texts, many conversion errors needed to be edited. This process was time-consuming. We spent around 100 hours at this stage for 25 English texts and 25 Thai translations. After that, original texts and translations were semi-automatically aligned, as we did not have at our disposal a reliable process for fully automatically aligning English and Thai texts. We spent around 40 hours at this stage. Thai has no delimiter to mark the end of sentences. As such, we aligned the section of the texts based on paragraphs in the English source texts instead of sentences. We used Microsoft Word as the editing interface and used the align structure in Sketch Engine (see <https://www.sketchengine.eu/user-guide/user-manual/corpora/setting-up-parallel-corpora/#tab-id-2>) for this purpose. Once the alignment was complete, the texts were ready to be loaded into Sketch Engine.

Before loading the aligned text we were keen to source a Thai tokenizer and part-of-speech tagger. We contacted Sketch Engine and we were advised that the default tokenization performed by Sketch Engine was sufficient and we were not required to tokenize the input ourselves. Nevertheless, in the tests we conducted we could search on the English side of our parallel corpus (and get parallel segments in Thai), but we could not do word-based searches on the Thai side. Therefore, we realized that we would have to find a tokenizer for the Thai texts after all.

We inquired of Sketch Engine again if they had any plan to integrate the SWATH tokenizer into Sketch Engine in the near future. Sketch Engine replied that they were uncertain if they would include the tokenizer soon; however, they suggested that there is a Windows binary file and a source code that can be compiled under most systems, and the best way would be to compile the program from the source code in Linux. We then started using the Linux operating system to tokenize our Thai texts and this task was very time-consuming. Fortunately, not so long after that we were contacted again by Sketch Engine to say that they had integrated SWATH into Sketch Engine. With their support, we did not have to tokenize the texts ourselves before loading them into the tool.

Regarding the part-of-speech tagger, upon ‘compilation’ of the English side of the corpus, Sketch Engine allows the user to specify a part-of-speech tagger that will be used to assign word class tags (for different forms of ‘noun’, ‘verb’, etc.) to individual word tokens in the corpus. We selected the default modified English Tree Tagger for this purpose. However, there is no comparable part-of-speech tagger available for Thai in Sketch Engine, and our efforts to source a part-of-speech tagger for Thai that could be implemented in Sketch Engine were not successful. The Thai side of our parallel corpus thus remains ‘untagged’.

It is easier to build a Thai monolingual corpus than an English-Thai parallel corpus. After converting the files into Text Only format, we then checked the conversion errors and fixed them. We spent around 125 hours for this stage. After that, we could load our Thai monolingual texts into Sketch Engine.

### Corpus Processing

After we had constructed the corpus, we also experienced some problems when we conducted search queries on the Thai side. Although Thai texts loaded into Sketch Engine are now automatically tokenized, they are not 100% percent correctly tokenized. This is not the fault of the tokenizer. Due to the non-orthographic nature of Thai words, it is hard to obtain 100% correct results even when Thai words are manually tokenized. Therefore, when we search for compound words containing two or more lexemes we have to search every possible way. For example, when we search for occurrences of

อย่างไรก็ตาม /yang/+/rai/+/ko/+/tam/ ‘however’, we need to try searching for อย่างไรก็ตาม, อย่างไร ก็ ตาม, อย่างไร ก็ตาม, and อย่างไรก็ ตาม to include all occurrences of the word.

In conclusion, it is not easy to compile a Thai corpus; however, it is also not impossible. At the time of writing this thesis, the major obstacle in compiling and processing Thai corpora is dealing with technical issues. If more computational assistance from experts in the area becomes available, and the functionality of corpus tools for Thai improves, we will be able to build Thai corpora much more easily and quicker, and obtain more accurate results, which will directly benefit research in Thai corpus linguistics in the future.