

Semantic prosody in Thai

Pornthip Supanfai

November 2016

(revised version March 2017)

Thesis submitted for the degree of PhD

Department of Linguistics and English Language

Lancaster University

Abstract

Semantic prosody is an important concept and has become a primary research interest in corpus linguistics. This thesis undertakes the groundwork of fundamental research into semantic prosody in Thai, a language which has not been subject to studies of semantic prosody before, to set out the parameters for subsequent research in this area.

In particular, it addresses these three research questions:

1. What are the advantages and disadvantages of the major approaches to semantic prosody proposed in the literature for describing semantic prosody in Thai?
2. What variation in semantic prosodies across genres can be identified for Thai words?
3. To what extent are the semantic prosodies of words identified as translation-equivalents in widely-used bilingual dictionaries in Thai and English similar or different?

The datasets employed in the analysis are the Thai National Corpus and the British National Corpus. To address each research question, a small number of Thai words are selected for the analysis. Two primary approaches, the polarity-oriented approach and the EUM-oriented approach, are employed to identify semantic prosody. Within the polarity-oriented approach, which is founded in work by Louw, Stubbs, and Partington, semantic prosody is identified based on collocates, and is restricted to the positive vs. negative opposition. Within the EUM-oriented approach, which is based in the studies of Sinclair, semantic prosody is identified by examining concordance lines for a pragmatic function or meaning that is spread across an extended unit of meaning.

The results of the analysis show that the two primary approaches to semantic prosody do operate successfully with the Thai data. A range of semantic prosodies are identified for /kreençay/ ‘considerate’, /kððhâykə̀ət/ ‘cause’, and /chôp/ ‘like’, the objects under study, by the two approaches. The discussion of these semantic prosodies

shows that the two approaches are useful for different purposes. The polarity-oriented approach is useful when one's aim is to investigate a word's tendency to co-occur with positive or negative words. Particularly, it reveals the hidden evaluative potential of words whose evaluation is not obvious from their core semantics. The EUM-oriented approach is, by contrast, suitable for the examination of an extended unit of meaning and its pragmatic function in the Sinclairian sense. They both also have some advantages and disadvantages in terms of practicality. On the issue of variation in semantic prosodies across genres, some variation is indeed found to exist. From the concordance analysis of 19 verbs, each in four different genres, namely academic writing, fiction, newspaper stories, and non-academic non-fiction, 21 different extended units of meaning are identified from 14 of the verbs. The level of variation in the use of these extended units of meaning across genres, which implies variation in semantic prosodies, is considerable with some extended units of meaning, but is limited with others. In particular, a notable contrast is identified between academic and fiction genres in terms of which extended units (and semantic prosodies) are common. Finally, the majority of the translation-equivalent pairs under study (36 out of 48) show the same semantic prosody; of these, most present a neutral semantic prosody. In cases where the pairs show different semantic prosodies, there are not any cases where one word in the pair shows a positive semantic prosody, and the other shows a negative semantic prosody, and vice versa. It is thus arguable that there is a relationship between semantic prosody in Thai and English – not a genetic or areal relationship, but one that arises from a functional basis, that is, the meanings that the pairs of words under study express in both languages.

Acknowledgements

First of all, I would like to express my deepest gratitude to my supervisor, Dr Andrew Hardie, for his expert guidance, invaluable comments, moral support, understanding, and patience with me throughout my PhD. Without Dr Hardie, this thesis would never have come into existence.

I would also like to thank you Dr Richard Xiao for believing in me and giving me an opportunity to pursue my PhD at Lancaster University.

I am very grateful to my external examiner, Professor Alan Partington, and my internal examiner, Professor Paul Baker, for their insightful comments, which helped make my thesis become much clearer and stronger.

My sincere thanks also go to Associate Professor Wirote Aroonmanakun, for granting me the permission to use the Thai National Corpus in this study, Professor Tony McEnergy, Associate Professor Prima Mallikamas, and Assistant Professor Raksangob Wijitsopon, for their useful comments on my thesis.

I would also like to thank you all my participants in my inter-rater agreement study.

My special thanks go to all my friends in Lancaster and the UK, especially P Jacky, P Pui, P Kwang, and N Mew, for being by my side when I was down, P Big, P Thom, and everyone at Siam Balcony. I would have been very lonely without your company.

I would also like to thank my best friends back in Thailand for being by my side during my ups and downs.

I am indebted to the Office of the Higher Education Commission and Mahasarakham University for sponsoring my study, which enabled me to pursue my dream and have a very comfortable life in the UK.

Finally, I would like to thank my mom, my dad, and my sister, Dobby, for their unconditional love, moral support, encouragement, and understanding. It was their love and support that encouraged me to move forward and finally complete this thesis.

Declaration

I declare that the work presented in this thesis is my own work, and has never been submitted, either in whole or in part, for the award of a higher degree elsewhere.

Table of Contents

1 Chapter 1 – Introduction	20
1.1 Context and objectives of research	20
1.2 Research questions	22
1.3 Introduction to the Thai language	24
1.3.1 Where is Thai spoken?	24
1.3.2 Genetic affiliation	25
1.3.3 A brief overview of the structure of Thai	26
1.3.3.1 Thai vocabulary	26
1.3.3.2 Syntax and morphology	27
1.3.3.3 Word order	39
1.4 Overview of the thesis	40
2 Chapter 2 – Literature review	43
2.1 Introduction	43
2.2 Neo-Firthian corpus linguistics	43
2.2.1 Introduction to neo-Firthian corpus linguistics	43
2.2.2 Open-choice principle vs. idiom principle	45
2.2.3 Collocation	48
2.2.3.1 Introduction to collocation	48
2.2.3.2 The Sinclairian or neo-Firthian approach to collocation	50
2.3 Approaches to semantic prosody	53
2.3.1 Louw’s approach to semantic prosody	53

2.3.1.1 Louw's introduction of semantic prosody	53
2.3.1.2 Louw's binary claim about semantic prosody	57
2.3.1.3 Whitsitt's criticism of Louw's characterisation of semantic prosody	59
2.3.1.4 Hunston's criticism of Louw's binary claim about semantic prosody	61
2.3.2 Sinclair's approach to semantic prosody	63
2.3.2.1 Sinclair's introduction of semantic prosody	63
2.3.2.2 Hunston's support for Sinclair's view of semantic prosody	68
2.3.3 Stubbs' approach to semantic prosody	71
2.3.3.1 Stubbs' early study of semantic prosody	71
2.3.3.2 Stubbs' later study of semantic prosody	73
2.3.4 Partington's approach to semantic prosody	76
2.3.4.1 Partington's early study of semantic prosody	76
2.3.4.2 Hunston's criticism of Partington's conceptualisation of semantic prosody and Morley and Partington's response	80
2.3.4.3 Partington's later study of semantic prosody	84
2.4 Lexical priming and semantic prosody	88
2.4.1 Introduction to lexical priming	88
2.4.2 What lexical priming adds to semantic prosody	91
2.5 Discussion of similarities and differences among the approaches to semantic prosody	94

2.5.1 Similarities among the approaches to semantic prosody	94
2.5.2 Differences among the approaches to semantic prosody	97
2.5.2.1 How semantic prosody is identified	98
2.5.2.2 Whether semantic prosody is only positive or negative	103
2.5.2.3 Where semantic prosody resides	108
2.5.2.4 Whether diachronic or synchronic considerations are involved in the concept of semantic prosody	110
2.5.3 Summary of similarities and differences among the approaches to semantic prosody	112
2.6 Studies of semantic prosody in English	116
2.7 Studies of semantic prosody in other languages	118
2.7.1 Berber Sardinha	119
2.7.2 Munday	120
2.7.3 Dam-Jensen and Zethsen	121
2.7.4 Xiao and McEnery	122
2.7.5 Wei and Li	123
2.7.6 Ebeling	125
2.7.7 Tognini-Bonelli	127
2.7.8 Summary of studies of semantic prosody in other languages	131
2.8 Concluding remarks	132
3 Chapter 3 – Methodology	134
3.1 Introduction	134
3.2 The Thai National Corpus	135

3.2.1 Structure of the TNC	136
3.2.2 Text Sampling	139
3.2.3 Encoding the TNC	139
3.2.4 The TNC at present	139
3.3 The British National Corpus	140
3.3.1 Structure of the BNC	140
3.3.2 Text sampling	142
3.3.3 Encoding and annotating the BNC	142
3.4 Methodology	142
3.4.1 Louw, Stubbs, and Partington's approach (the polarity-oriented approach)	142
3.4.1.1 General considerations	142
3.4.1.2 Type vs. token frequency	147
3.4.1.3 Inter-rater reliability	150
3.4.2 Sinclair's approach (the EUM-oriented approach)	155
3.5 Concluding remarks	157
4 Chapter 4 – Comparison of the polarity-oriented approach and the EUM-oriented approach	159
4.1 Introduction	159
4.2 Approaches to the analysis	159
4.2.1 Louw, Stubbs, and Partington's approach (the polarity-oriented approach)	159
4.2.2 Sinclair's approach (the EUM-oriented approach)	161

4.3 Selection of Thai words	162
4.4 Results	163
4.4.1 Louw, Stubbs, and Partington’s approach	163
4.4.1.1 /kreɛŋcay/ ‘considerate’	163
4.4.1.2 /kə̀əhâykə̀ət/ ‘cause’	163
4.4.1.3 /chôp/ ‘like’	165
4.4.1.4 Discussion	166
4.4.2 Sinclair’s approach	169
4.4.2.1 /kreɛŋcay/ ‘considerate’	169
4.4.2.2 /kə̀əhâykə̀ət/ ‘cause’	177
4.4.2.3 /chôp/ ‘like’	181
4.4.2.4 Discussion	186
4.5 General discussion	187
4.6 Concluding remarks	194
5 Chapter 5 – Variation in semantic prosodies across genres	195
5.1 Introduction	195
5.2 Approach to the analysis	195
5.3 Genres in the TNC	196
5.4 Selection of Thai words	198
5.5 Results	201
5.5.1 /pràkàat/ ‘announce, declare’	201
5.5.2 /càp/ ‘touch, arrest, etc.’	205
5.5.3 /klaaypen/ ‘become’	210

5.5.4 /hǎn/ ‘turn’	212
5.5.5 /ʔaasǎy/ ‘live, rely on, refer to’	214
5.5.6 /yɔɔm/ ‘resolve’	217
5.5.7 /yók/ ‘lift, refer, move, give, dismiss, draft (a document)’	218
5.5.8 /yũun/ ‘stand’	223
5.5.9 /khûapkhum/ ‘control, supervise, confine’	226
5.5.10 /níyom/ ‘like’	230
5.5.11 /waan/ ‘put, be placed, set, walk’	231
5.5.12 /rɔɔ/ ‘wait, suspend’	234
5.5.13 /ʔànúyâat/ ‘allow’	236
5.5.14 /sǎɔn/ ‘teach’	240
5.5.15 /càtkaan/ ‘manage, deal with’	242
5.5.16 /hǎay/ ‘recover, disappear’	245
5.5.17 /bankháp/ ‘force, execute, steer’	248
5.5.18 /duulɛɛ/ ‘take care of’	253
5.5.19 /còp/ ‘graduate, end/finish’	256
5.6 Discussion	260
5.6.1 Single-word units of meaning	260
5.6.2 Extended units of meaning	264
5.6.3 Extended units of meaning reconsidered from the polarity-oriented perspective	268
5.6.4 Variation in use of single-word units of meaning	278
5.6.5 Variation in use of extended units of meaning and	278

semantic prosodies	
5.7 Concluding remarks	285
6 Chapter 6 – Analysis of semantic prosodies of translation-equivalents	286
6.1 Introduction	286
6.2 Approach to the analysis	286
6.3 Prosodic strength	287
6.4 Selection of translation-equivalent pairs	290
6.4.1 Selection of Thai words	290
6.4.2 Selection of English translation-equivalents	291
6.5 Results	296
6.5.1 Lists of the collocates of each Thai word and its English translation-equivalent	296
6.5.1.1 /phõnɲaan/ vs. <i>achievement</i>	296
6.5.1.2 /thõɲthìn/ vs. <i>district</i>	297
6.5.1.3 /pràwàtsàat/ vs. <i>history</i>	298
6.5.1.4 /thoorásàp/ vs. <i>telephone</i>	300
6.5.1.5 /phûukhon/ vs. <i>people</i>	301
6.5.1.6 /phâappháyon/ vs. <i>film</i>	302
6.5.1.7 /pràsithíphâap/ vs. <i>efficiency</i>	303
6.5.1.8 /ɲâankhăy/ vs. <i>condition</i>	304
6.5.1.9 /phûuyày/ vs. <i>adult</i>	305
6.5.1.10 /sùkkhâphâap/ vs. <i>health</i>	306
6.5.1.11 /rátthàthammánuun/ vs. <i>constitution</i>	307

6.5.1.12 /pâwmăay/ vs. <i>target</i>	308
6.5.1.13 /chaaynùm/ vs. <i>young man</i>	309
6.5.1.14 /siŋwêetlóm/ vs. <i>environment</i>	310
6.5.1.15 /pràsòpkaan/ vs. <i>experience</i>	311
6.5.1.16 /khwaamtôŋkaan/ vs. <i>desire</i>	312
6.5.1.17 /tàwantòk/ vs. <i>west</i>	313
6.5.1.18 /ʔitthíphon/ vs. <i>influence</i>	314
6.5.1.19 /chûamooŋ/ vs. <i>hour</i>	315
6.5.1.20 /phúunthăan/ vs. <i>foundation</i>	316
6.5.1.21 /sàthăanâkaan/ vs. <i>situation</i>	317
6.5.1.22 /khamtòp/ vs. <i>answer</i>	318
6.5.1.23 /raayláʔiat/ vs. <i>detail</i>	319
6.5.1.24 /phômêe/ vs. <i>parents</i>	320
6.5.1.25 /nêewkhít/ vs. <i>idea</i>	320
6.5.1.26 /ʔèkkàchon/ vs. <i>individual</i>	321
6.5.1.27 /naayókrátthàmontrii/ vs. <i>prime minister</i>	322
6.5.1.28 /ʔèkkàsăan/ vs. <i>document</i>	325
6.5.1.29 /nêewthaŋ/ vs. <i>way</i>	326
6.5.1.30 /muanthay/ vs. <i>Thailand</i>	327
6.5.1.31 /lâkthăan/ vs. <i>evidence</i>	328
6.5.1.32 /sápsĭn/ vs. <i>property</i>	329
6.5.1.33 /khamsàŋ/ vs. <i>order</i>	330
6.5.1.34 /ŋóppràmaan/ vs. <i>budget</i>	332

6.5.1.35 /náamman/ vs. <i>oil</i>	333
6.5.1.36 /khâarâatchákaan/ vs. <i>civil servant</i>	335
6.5.1.37 /ʔoŋkoon/ vs. <i>company</i>	336
6.5.1.38 /cìtcay/ vs. <i>mind</i>	338
6.5.1.39 /mâattràthhãan/ vs. <i>standard</i>	339
6.5.1.40 /nákrian/ vs. <i>student</i>	340
6.5.1.41 /khwaamsãamâat/ vs. <i>ability</i>	341
6.5.1.42 /mùubâan/ vs. <i>village</i>	342
6.5.1.43 /sàthhãanthhîi/ vs. <i>place</i>	343
6.5.1.44 /chaawbâan/ vs. <i>villager</i>	344
6.5.1.45 /kitcàkam/ vs. <i>activity</i>	345
6.5.1.46 /bòtbàat/ vs. <i>role</i>	346
6.5.1.47 /khunnáphâap/ vs. <i>quality</i>	347
6.5.1.48 /kitcàkaan/ vs. <i>business</i>	348
6.5.2 Methodological issues	350
6.5.3 Semantic prosody of each Thai word and its English translation-equivalent	360
6.6 Discussion	363
6.7 Concluding remarks	369
7 Chapter 7 – Conclusion	371
7.1 Introduction	371
7.2 Findings of the study	371
7.3 Limitations of the study	378

7.4 Contributions	383
7.5 Possible future research	385
7.6 Final remarks	387
References	388
Dictionaries	391
Appendix – 1 Phonology and writing system	392
Appendix – 2 The questionnaire used in the inter-rater agreement study	397
Appendix – 3 List of abbreviations used in line 2 of Thai examples	398

List of Tables

Table 1.1	Aspect markers in Thai	32
Table 1.2	Pragmatic particles in common use	38
Table 2.1	Words/Expressions studied by Louw and their prosodies	55
Table 2.2	Differences among the four scholars' approaches to semantic prosody	113
Table 2.3	Examples of semantic prosodies in English	116
Table 3.1	Summary of weights of domain, medium, and time in the TNC	137
Table 3.2	An overview of the genre categorisation used in the TNC	138
Table 3.3	Summary of weights of domain, medium, and time in the BNC	141
Table 3.4	Level of agreement among participants	151
Table 4.1	Notations for extended units of meaning	161
Table 4.2	Frequency of occurrence of each extended unit of meaning containing /kreenɕay/	176
Table 4.3	Distribution of co-occurring noun and verb tokens across meaning categories	181
Table 5.1	Selected verbs, their meanings, and their frequency of occurrence	200
Table 5.2	Extended units of meaning and their semantic prosodies	266
Table 5.3	Extended units of meaning, their semantic prosodies and evaluation	273
Table 5.4	Extended units of meaning and their frequency of occurrence across genres	280

Table 6.1	Selected Thai words, their English translation-equivalents, their frequency, and their frequency per million words	293
Table 6.2	The number of positive, negative, and neutral collocates, semantic prosody, and prosodic strength of each Thai word and its English translation-equivalent	360
Table A1.1	Initial consonants	392
Table A1.2	Final consonants	392
Table A1.3	Initial consonant clusters	393
Table A1.4	Vowels	393
Table A1.5	Diphthongs	393
Table A1.6	Five different tones in Thai	394
Table A1.7	Thai consonant letters chart	395
Table A1.8	Vowel letters and the sound they represent	396

List of Figures

Figure 1.1	Tai-Kadai language family	26
Figure 2.1	An example of a phrase structure tree diagram	46
Figure 2.2	An extended unit of meaning whose core is <i>naked eye</i>	68

Chapter 1 – Introduction

The aim of this chapter is to introduce the thesis. This will involve introducing my research objectives (in section 1.1) and my research questions (in section 1.2). I will also provide a brief overview of some of the main features of the Thai language (in section 1.3), as this is a necessary background to understand the materials that will follow later in the thesis. Finally in this chapter (in section 1.4), I will present the organisation of the thesis.

1.1 Context and objectives of research

Semantic prosody has become an important concept in corpus linguistics (Bednarek 2008: 119; Whitsitt 2005: 283;), and it has attracted much interest in the past 15 to 20 years (Ebeling 2014: 161; Stewart 2010: 6). Because it is a relatively new concept, there is no consensus on its definition (Zhang 2009: 2), and even its name is controversial (Partington 2014: 279). As well as *semantic prosody*, which, according to Louw (1993: 158), was coined by John Sinclair, and which seems to be the most widely used terminology, at least four other names have been used to refer to the concept. Stubbs (2001: 65-66; see also Partington 2014: 279) suggests that it might be better to use the terms *pragmatic prosody* or *discourse prosody*. Bublitz (2003 cited in Partington 2014: 279) suggests the term *emotive prosody*. Most recently, Partington (2014: 279) explicitly argues that “the most descriptive denomination would be *evaluative prosody*.” However, despite this considerable debate on the notion and name of semantic prosody, there are primarily four scholars – namely Louw (1993), Sinclair (2004), Stubbs (1995; 2001), and

Partington (1998; 2004; 2014) – who have substantially contributed to the definition of the concept.

Louw was the first person to introduce the term semantic prosody to the public, although he credits Sinclair for having provided him with the term and the concept (Whitsitt 2005: 283-186). Louw (1993: 159) defines semantic prosody as “a consistent aura of meaning with which a form is imbued by its collocates”. In particular, a lexical item may be said to display either a *positive* or *negative* semantic prosody, depending on the context it habitually occurs in. However, Louw’s definition of semantic prosody is quite different from those of Sinclair, Stubbs, and Partington, who also differ from each other (see section 2.3 and section 2.5). As well as these four scholars, many other scholars have explored semantic prosody and applied the concept in a variety of areas, e.g. Baker (2006) and Mautner (2007) (see section 2.6).

The four scholars named above all explore semantic prosody in English. However, research on semantic prosody is not restricted to the English language, even though studies of the phenomenon in other languages are still few in number. To my knowledge, semantic prosody has also been explored in Spanish (Munday 2013), Italian (Tognini-Bonelli 2001), Portuguese (Berber-Sardinha 2000), Danish (Dam-Jensen and Zethsen 2008), Norwegian (Ebeling 2013; 2014) and Chinese (Xiao and McEnery 2006) (see section 2.7). These studies often feature comparison of semantic prosodies between English and the language in question.

We thus see that the number of languages where semantic prosody has been explored is still limited. Moreover, the studies of semantic prosody in languages other than English seem in large part to be restricted to contrastive studies between those

languages and English. This thesis aims to further this field of research; its purpose is to examine semantic prosody in Thai. There are three reasons why Thai is a suitable language to examine in this context. First, semantic prosody in this language has never before been studied. Second, it is genetically, and thus lexically, unrelated to English. (There are some loanwords in Thai that are shared with English. However, this thesis will not look at those items.) Third, Thai is unrelated to Chinese, the only East Asian language that has been studied so far, as noted above. As mentioned earlier, studies of semantic prosody in languages other than English have largely been contrastive studies. This thesis does aim to compare and contrast semantic prosody between Thai and English. However, this contrastive study will only be *part* of the thesis; the study also aims to look at semantic prosody in Thai in its own terms. This is an important aspect of the novelty of the thesis.

1.2 Research questions

The study aims to carry out the groundwork of fundamental research into semantic prosody in Thai to set out the parameters for subsequent research in this area. This will be accomplished by looking at the following three specific questions:

1. What are the advantages and disadvantages of the major approaches to semantic prosody proposed in the literature for describing semantic prosody in Thai?
2. What variation in semantic prosodies across genres can be identified for Thai words?

3. To what extent are the semantic prosodies of words identified as translation-equivalents in widely-used bilingual dictionaries in Thai and English similar or different?

The theoretical background assumed by these questions will be outlined in Chapter 2; moreover, the study will adopt a corpus-based methodology as explained in Chapter 3. A small number of relevant Thai words will be specifically selected for each question. To address research question 1, each word will be examined using two different approaches. The first approach will be based on Louw, Stubbs and Partington's approaches (which are fairly similar). This approach will primarily look at individual co-occurring items around the Thai words under study. The second approach will be based on Sinclair's approach to semantic prosody. Each word will be analysed for its semantic prosody within the framework of Sinclair's model of the extended unit of meaning (see section 3.4). The approach to be adopted to address research question 2 and research question 3 will be determined on the basis of the results obtained from research question 1. That is, the effectiveness of the two approaches as well as issues relating to practicality will be taken into account when choosing the optimal approach for the remaining two research questions. The corpus to be used in the research is the Thai National Corpus (TNC) and, for contrastive analysis, the British National Corpus (BNC) (see section 3.2 and section 3.3).

It is hoped that the results of this study, which is one of the very first studies to examine Thai using a corpus-based approach, will enhance our understanding of the Thai language. Moreover, the fact that this study will be the first study to adopt both of the two primary approaches to semantic prosody within a single analysis will hopefully provide

us with a deeper comprehension of the differences between the two approaches and what each approach is useful for. In addition, an in-depth comprehension of semantic prosody in Thai will pave the way for much new research on stylistics, (critical) discourse analysis, sociolinguistics, language teaching, and linguistic theory - each of which has, in English linguistics, been informed by the study of semantic prosody - in the Thai language. Furthermore, the study of the semantic prosodies of translation-equivalents will provide invaluable insights for translators as well as language learners. Most importantly, the study will function as a test of the cross-linguistic adequacy of existing linguistic theories in the area of semantic prosody, thus pushing the boundaries of linguistic knowledge. Looking at Thai will help to broaden our understanding of semantic prosody generally – and to move forwards to answering the greater question of whether it works in a similar way across *all* languages.

1.3 Introduction to the Thai language

1.3.1 Where is Thai spoken?

Thai, formerly known as Siamese, is the national and official language of Thailand (Smyth 2002: 1), a Southeast Asian country with a population of 65,729,098 as of the year 2015 (Department of Provincial Administration). Four distinct dialects of Thai are spoken in different regions of the country: the Northern dialect (Kam Muang, Lan Na or Yuan), the Northeastern dialect (Isarn or Lao), the Southern dialect, and the Central dialect (Iwasaki and Ingkaphirom 2005: 1). Standard Thai, the object of this study, is one form of the Central dialect.

Standard Thai is used by educated, middle class people in the central region of Thailand, especially in Bangkok, the capital. It is the native language of 20 to 25 million people (Iwasaki and Ingkaphirom 2005: 1). However, as Standard Thai is taught in schools and used in the mass media, it is extensively used and understood throughout the country.

1.3.2 Genetic affiliation

Thai is a language of the Tai family (Campbell and King 2013: 1663; Smyth 2002: 1). It is classified as a member of the Southwestern branch of Tai, along with languages spoken in Laos, northern Vietnam, Myanmar, India, and southern China (Iwasaki and Ingkaphirom 2005: 1). The other two branches of Tai are the Central branch, which is comprised of languages spoken in northern Vietnam and southern China such as Nùng and Tày, and the Northern branch, which includes other languages of southern China, such as Bouyei and Seak. The Tai language family is a subgroup of the Kam-Tai language family, which belongs to the larger Kadai or Tai-Kadai language family. Figure 1.1 illustrates the Kadai language family, as proposed by Edmondson and Solnit (1997 cited in Iwasaki and Ingkaphirom 2005: 2).

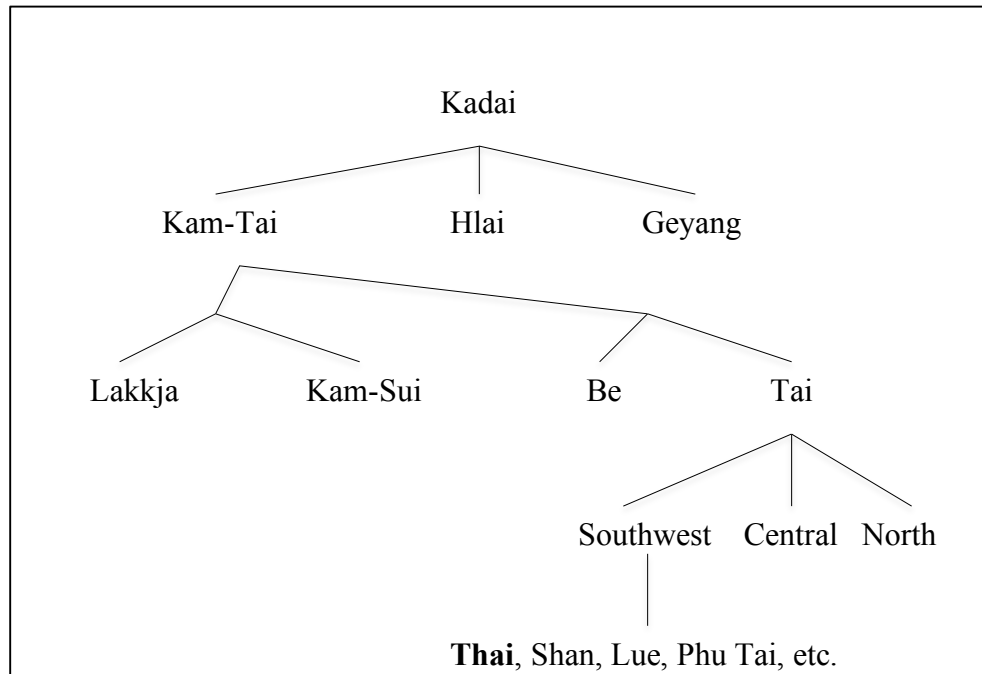


Figure 1.1 Tai-Kadai language family (after Iwasaki and Ingkaphirom 2005: 2)

1.3.3 A brief overview of the structure of Thai

The following cursory discussion of the structure of the Thai language, including all the linguistic examples (all of which are given in phonetic transcription rather than Thai script), is drawn primarily from Iwasaki and Ingkaphirom (2005), where not otherwise noted. The phonetic transcription is a modified IPA, as used by those authors. An overview of the phonology and writing system of Thai, which are less directly relevant to the purposes of this thesis than Thai grammar and lexis, is provided in Appendix 1.

1.3.3.1 Thai vocabulary

Most of the Thai lexicon consists of loanwords. Out of approximately 20,000 words listed in the *Royal Institute's Dictionary*, only a few thousand are not loanwords.

Most of the original Thai words are monosyllabic (Smyth 2002: 1), such as /fan/ ‘tooth’, /pay/ ‘go’, /kin/ ‘eat’, /dam/ ‘black’, /phèt/ ‘spicy hot’. A large number of words have been borrowed from Sanskrit and Pali, the languages of Buddhism, e.g. /bun/ ‘virtue, merit’, /manút/ ‘human’, /phayaayaam/ ‘try’, and /kròot/ ‘angry’. There are also Chinese and Khmer loans. The Chinese loans include numbers, such as /nə̀ŋ/ ‘one’, /sǒŋ/ ‘two’, and /sǎam/ ‘three’; body parts, such as /hǔa/ ‘head’, /ʔeew/ ‘waist’, and /khǎa/ ‘leg’; animals, e.g. /máa/ ‘horse’, /kày/ ‘chicken’, /cháaŋ/ ‘elephant’; and many more. Khmer loans include /saphaan/ ‘bridge’, /talòk/ ‘comedy’, /kraasíp/ ‘whisper’, /chalàat/ ‘clever’, and /sanùk/ ‘fun’, to name but a few. There are also recent loans from English and other languages.

1.3.3.2 Syntax and morphology

Thai is an isolating language; Thai verbs and nouns do not inflect for number or gender, for example. Thai is a classifier language, in that different classifiers are used when quantifying different kinds of nouns.

Iwasaki and Ingkaphirom (2005: 8) categorise Thai words into 14 word classes, although they also note that there is no consensus on this point. Some Thai words can be classified into more than one word class. These words are undergoing the process of grammaticalisation, in which content words, such as nouns and verbs, change to function words, such as prepositions and auxiliaries.

Let us begin with words related to nouns, and first the noun themselves. Thai has both simple and compound nouns. Most compound nouns have a noun as the first element (the head) and a noun, verb, or adjective as the second component. Thai nouns

can function as a subject, an object of a verb or preposition, or a predicate nominal after the copula /pen/ or /khǘ/. They can be modified by numerals, demonstratives, interrogative and indefinite determiners, genitives, adjectives, prepositional phrase modifiers, and relative clauses, all of which appear after the head noun.

As well as nouns, there exist a great number of pronouns in Thai, including nine first person- pronouns, eight second-person pronouns, and five third-person pronouns in common use in everyday language. The choice of pronoun is based on sociolinguistic criteria, such as the level of formality and the referent's sex. Thai pronouns can function as a subject, or an object of a verb or preposition. They can be modified by demonstratives, but not by adjectives, numeral phrases, genitive phrases, or relative clauses.

Related to pronouns are demonstratives. Thai demonstratives distinguish between the proximate /ní/ (close to the speaker), medial /nán/ (away from the speaker), and distal /nóon/ (further away from the speaker). These forms are used as specifiers of nouns; the corresponding independent pronouns are /nî/, /nân/, and /nôn/ respectively.

Thai marks spatial, temporal, and conceptual relationships among nouns and pronouns via prepositions. Thai prepositions are immediately followed by their object noun phrase. Some examples are /nay/ 'in', /bon/ 'on, on top of', and /rim/ 'on the edge of'.

An important aspect of the noun phrase in Thai is the numerals. The Chinese number system has been borrowed in Thai. The cardinal numerals from one to ten are: /nǎn/, /sǒŋ/, /sǎam/, /sìi/, /hâa/, /hòk/, /cèt/, /pèt/, /kâw/, and /sìp/.

Unlike English, Thai has a noun classifier system. Many Thai nouns need classifiers when they are modified by a numeral. A numeric phrase must consist of a number and a classifier. For example, ‘three houses’ would be expressed as *house + three + classifier* (here, /lǎŋ/). There are 30-40 classifiers in common use.

Moving on to words related to verbs, let us begin with the verbs themselves. Thai verbs differ from English verbs in a number of ways. Unlike English verbs, Thai verbs do not inflect for person, number, and tense. However, like English verbs, Thai verbs can mostly be classified as transitive, intransitive, or ditransitive.

The use of serial verb structures, where two or more verbs or verb phrases are used in sequence without any coordinator or subordinator, is common in Thai. There are three types of serial verb structure in Thai: the ‘subordination’ type, the ‘coordination’ type, and the ‘hybrid’ type.

In the subordination type, a verb of ‘desire’, ‘willingness’, or ‘intention’ appears as the main verb, and is followed by a subordinate verb. This structure is similar to that of the English sentence *I tried to sleep*, for example, where the main verb *tried* subordinates the infinitive *sleep*. In Thai, there may be a marker /cà/ between the main verb and the subordinate verb.

In the coordination type, two or more verbs or verb phrases are placed together as if they were linked by *and* – but with no explicit coordinator. There are three sub-types of coordination serialisation: sequential/purposive serialisation, motion serialisation, and simultaneous serialisation. In the sequential/purposive serialisation and the motion serialisation, two or more consecutive activities happen one after another. In the simultaneous serialisation, two or more actions happen at the same time. In the

subordination type and the coordination type, the verbs must share the same subject. In the hybrid type, however, the two verbs may have different subjects. There are two subtypes of hybrid serialisation: causative serialisation and resultative serialisation.

The causative serialisation “depicts an event consisting of a causing and resulting situation” (Iwasaki and Ingkaphirom 2005: 239). One of the typical structures of this subtype of serialisation is formed by the causative markers, /hây/, /tham/, or /tham-hây/, as in Example 1.1¹. The term *causative serialisation* is specific to Thai grammar, and is equivalent to the general term *periphrastic causative*.

Example 1.1

somchaay	sàŋ	lûuk	(hây)	pay	rîak	mǔɔ
(name)	order	child	(give/CAUS)	go	call	doctor
‘Somchai ordered his child to call for a doctor.’						

Similarly to the causative serialisation, the resultative serialisation also depicts a causing and resulting situation. But in contrast to the causative serialisation, the resulting situation “represents a natural and expected consequence of the initial causing event” (Iwasaki and Ingkaphirom 2005: 239), as in Example 1.2.

Example 1.2

phîi	yók	krapǎw	khûn
OS	lift	bag	ascend/DIR
‘M brother lifted the bag up’			

Among the grammatical elements used alongside verbs are negators. There are four negators in Thai: /mây/, /mây-chây/, /yàa/, and /plàaw/. /mây/, a general negator, is placed before a verb, as in /mây pay/ ‘not go’; an adjective, as in /mây sǔay/ ‘not beautiful’; or certain auxiliary verbs, as in /mây khuan/ ‘should not’. /mây-chây/ appears

¹ The grammatical labels in the examples quoted from Iwasaki and Ingkaphirom (2005) and Smyth (2002) are as in the original and are dissimilar in some ways to my own system of glossing.

before a nominal predicate, e.g. /mây-chây khon thai/ ‘not a Thai person’. /yàa/ appears before a verb to form an imperative construction, such as /yàa pay/ ‘Don’t go’. Finally, /plàaw/ is used as a negative answer to some types of questions.

Many aspects of verbal grammar are marked via auxiliary verbs. Thai auxiliary verbs add grammatical features such as aspect, modality, potentiality, and direction to the main verb.

Thai has a rich aspect system. Aspect in Thai can be expressed by several words. These aspect markers may appear before the verb (pre-verbal aspectual auxiliaries), or after the verb (post-verbal aspectual auxiliaries); most post-verbal aspectual auxiliaries can also be used as a main verb. A single verb may have both pre- and post-aspect markers, as shown in Example 1.3.

Example 1.3

kháw	nâŋ	sǎ-cay	yùu	khon	diaw
3	sit/ASP	disappointed	stay/ASP	CLS	single

‘He is disappointed in solitude now.’

The verb /sǎ-cay/ has two aspect markers, /nâŋ/ and /yùu/, which together indicate the continuous aspect.

There are 20 aspect markers in Thai. These markers can be broadly categorised into four groups, as illustrated in Table 1.1. The three dots indicate the position of a lexical verb.

General categories	Aspect marker	Aspect indicated
Perfective	.../léɛw/	General perfective ('has done')
Imperfective	.../yùu/ /nâŋ/...(/yùu/ /khəəy/...(/yùu/ .../pay/ .../pay/.../pay/ .../?aw/.../?aw/ /kamləŋ cà/...(/yùu/ /cuan cà/ .../còp/; .../sèt/	Continuous (e.g. Progressive) ('is doing etc.) Special progressive ('is doing') Habitual (Prolonged activity) ('has been doing') Continuative ('keep doing') Continuative – simultaneous ('keep doing two actions') Continuative – intensive ('keep doing') Inceptive ('is about to do') Pre-inceptive ('is almost done') Terminative ('has finished doing')
Perfect/Anterior	.../léɛw/ .../maa/ .../pay/ .../sǎa/ .../wáy/ /phôŋ (cà)/... /khəəy/...	General perfective/anterior ('has finished doing') General perfective/anterior ('has finished doing') Completive ('has done') Completive ('has done') Purposive ('has done for future') Immediate perfect/anterior ('has just done') Experiential ('has ever done')
Change-of-state ('Inchoative')	/dây/... .../khân/ .../loŋ/	General changed situation Upward change Downward change

Table 1.1 Aspect markers in Thai (after Iwasaki and Ingkaphirom 2005: 149-150)

Thai modal auxiliaries express concepts such as obligation, degree of speaker confidence, and how the speaker obtained the information. The modal auxiliaries expressing obligation include /tôŋ/ 'must', /campen (cà) (tôŋ)/ 'be necessary', /khuan (cà)/ 'should', and /nâa (cà)/ 'should'. These auxiliaries precede the main verb in a clause. In a negative sentence, the negative marker /mây/ is placed before the auxiliary (see section 1.3.3.2). The negative marker may be placed either before or after /tôŋ/, however. While for /tôŋ mây/, the negation scope is over the main verb, for /mây tôŋ/,

the negation scope is over the modal auxiliary. The modal auxiliaries that describe the speaker's degree of confidence are, from the highest to the lowest degree of confidence, /tôŋ/; /khuan/ and /nâa/; and /khoŋ/ and /ʔaat/. Like those expressing obligation, these auxiliaries are all placed before the main verb. In a negative sentence, /mây/ is placed after /tôŋ/, /khoŋ/, and /ʔaat/, but before /nâa/. Special modal constructions are used to express the source of the speaker's information, e.g. /duu-măan (wâa)/ 'it looks like; it appears', /tham-thâa/ 'make it appear like', and /dâyyn wâa/ 'I hear that'.

Thai has three potential auxiliaries. These auxiliaries express different concepts such as possibility, ability, permission, and physical and mental readiness/strength. These different meanings are all covered in English by 'can'. Despite the difference in the concept each auxiliary expresses, they all appear after a clause and are negated by placing /mây/ immediately before them. They can also appear alone or with a verb in a positive answer and appear with /mây/ in a negative answer. The three potential auxiliaries are /dây/ 'can', /pen/ 'can', and /wăy/ 'can'.

The /dây/ potential auxiliary expresses the meanings of 'possibility', 'ability', and 'permission'. Example 1.4 expresses the meaning of 'possibility' of /dây/.

Example 1.4

phrûŋ-ní	chûay	maa	thîi	bâan	dây	máy
tomorrow	help	come	at	house	POT	QP

'Can you come to my house tomorrow?'

Due to the different meanings /dây/ can express, sometimes another expression is added to clarify the function of /dây/. /khǒo/ 'beg' or /khǒo ʔànúyâat/ 'I'd like to ask for permission' is added in permission questions, whereas the potential adverb /săamâat/ is added when ability needs to be specified.

The /pen/ potential auxiliary is restricted to the meaning of ‘ability acquired through practice’, as in Example 1.5.

Example 1.5

khon	thay	sùanyà	kin	plaa-dip	mây	pen
people	Thai	mostly	eat	fish-raw	NEG	POT

‘Most Thai cannot eat raw fish.’

The /wǎy/ potential auxiliary “indicates possession of a physical or mental condition to carry out activities,” (Iwasaki and Ingkaphirom 2005: 354) as in Example 1.6.

Example 1.6

dǎn	tò	wǎy	máy
walk	continue	POT	QP

‘Can (you) continue to walk?’

There are some directional auxiliary verbs in Thai, such as /pay/ ‘go’, /maa/ ‘come’, /khân/ ‘ascend’, /loŋ/ ‘descend’, /khâw/ ‘enter’, and /ʔòk/ ‘exit’. These auxiliaries indicate “the direction towards the center of discourse, e.g. the speaker” (Iwasaki and Ingkaphirom 2005: 18) when they follow a verb of motion, as in /klàp maa/ ‘return come’.

Moving on to modifying words, let us consider adjectives first. Thai adjectives are primarily used to modify nouns. Some examples of adjectives in Thai are /yà/ ‘big’, /klom/ ‘round’, and /lò/ ‘handsome’. When an adjective modifies a noun, it is placed after the noun it modifies, as in Example 1.7.

Example 1.7

rót	(khan)	phɛɛŋ
car	(CLS)	expensive

‘an expensive car’

Thai adjectives may also function as predicates in a sentence, as in Example 1.8 where /phɛɛŋ/ ‘expensive’ is the predicate of the sentence.

Example 1.8

rót	khan	níi	phɛɛŋ
car	CLS	this	expensive

‘This car is expensive.’

Some adjectives can be used as manner adverbs, such as /rew/ ‘fast’, /cháa/ ‘slow/slowly’, /sùphâap/ ‘polite/politely’. When they function as a manner adverb, they are placed immediately after the verb, as in Example 1.9.

Example 1.9

nũu	tôŋ	phũut	sùphâap	kàp	phũuyà	ná
mouse.2	must	speak	polite	with	adult	PP

‘You have to speak politely to adults.’

In comparative structures, a preposition phrase consisting of /kwàa/ ‘than’ followed by a noun is generally used, so the adjective itself does not change. In the comparative correlative construction, /yîŋ...yîŋ.../ is used, as in /yîŋ rew yîŋ dii/ ‘The faster, the better.’ This structure is not limited to adjectives and manner adverbs, but also operates with verb phrases. In superlative structures, /thîi-sùt/ ‘the most’ is generally placed after adjectives or manner adverbs.

Adjectives and manner adverbs can be intensified by emphatic reduplication or intensifiers. In the emphatic reduplication process, the first adjective has emphatic high tone, instead of the word’s normal tone, as in /wáan-wǎan/ ‘really sweet’ and /kɛŋ-kɛŋ/ ‘really well’. Adjectives and manner adverbs may also be modified by intensifiers. Some intensifiers are /mâak/ ‘very’, /caŋ/ ‘extremely, quite’, /nâa-duu/ ‘extremely, really’, and /thii-diaw/ ‘very, quite’. These intensifiers are placed after the word they modify, as in

Example 1.10.

Example 1.10

kháw	wîŋ	rew	mâak/caŋ
3	run	fast	very/quite
'He runs very fast.'			

As well as adjectives and manner adverbs, Thai has frequency and quantity adverbs, such as /samǎə/ 'always', /bòɔy/ 'often', and /râay/ 'all the time', /baaŋthii/ and /baaŋkhráŋ/ 'sometimes', /sùanmâak/ 'mostly', and /dooythûapay/ 'generally'.

Thai does not have one-word temporal and locational adverbs equivalent to English 'now' and 'here'. Temporal adverb phrases and locational adverb phrases are instead used to indicate time and place information respectively. Some examples are /wan-níi/ 'today' lit. 'this day', /wancan níi/ 'this Monday', /thîi bâan/ 'at home', and /nay tûuyen/ 'in the fridge'.

Finally let us consider the remaining grammatical categories, conjunctions and particles. There are a number of conjunctions in Thai. Some co-ordinate individual words, such as /kàp/ and /kà/ 'and', and /rǎw/ 'or'. Others function as clausal co-ordinating conjunctions, such as /lé/ 'and', /tèɛ/ 'but', and /rǎw/ 'or'. There are also subordinating conjunctions, including /welaa (thîi)/ 'when', /phɔɔ/ 'as soon as', /khanà thîi/ 'while', /thâa/ 'if', and /phró wâa/ 'because'. However, it is possible for the conjunction marking an adverbial subordinate clause to be omitted.

In addition to these conjunctions, Thai has main clause markers, which optionally indicate the main clause of a complex sentence. These markers appear before the verb phrase of the main clause. The four main clause markers are /kâ/, /thǎŋ/, /lǎəy/, and /cəŋ/.

As well as conjunctions, Thai has a great number of particles. Thai has five sub-types of particles: speech-level marking particles, question particles, pragmatic particles, one linking particle, and one modal particle.

Speech-level marking particles indicate the level of formality between the interlocutors. Many are gender-specific. For example, /khráp/ is exclusively used by males; /khà/ and /khá/ are used by females at the high level of formality. On the other hand, /wá/ and /wà/ are used by both males and females at the low formality level.

Question particles are added at the end of a statement to form yes/no questions and tag questions. The four different question particles are /mǎy/, /rǎu-plàaw/, /rǎu-yan/, and /rǎu/. To form tag questions, /chây-mǎy/, /chây-rá-plàaw/, or /mây-chây-rá/ is placed at the end of a statement.

The pragmatic particles are many in number and elusive in meaning. More than one pragmatic particle may appear in a sentence, and in such cases, these particles generally occur in a fixed sequence. Thai pragmatic particles can be broadly categorised into three groups, in accordance with the type of sentence where they can appear. Table 1.2 lists some pragmatic particles in common use.

General category	Pragmatic particles	Functions
General particles (appear with statement, question, and command)	/ná/ - /nà/ - /à/ /sí/ - /sî/ - /sì/ /læy/ /chiaw/ /sía/ - /sá/	Common ground Authoritative confirmation Emphasis Emphasis, surprise Perfect/anterior; mild encouragement
Information-oriented particles (appear with statement and question)	/lá/ - /lâ/ - /ʔà/ /lê/ - /ʔâ/ /lâ/ - /lá/ - /ʔà/ /ròk/ - /lòk/ - /dòk/ /ɲay/	Elaboration request Perfect/anterior; focus Conclusion Counterargument Directing attention
Action-oriented particle (appear only with command)	/thə/ - /hə/	Suggestion

Table 1.2 Pragmatic particles in common use (after Iwasaki and Ingkaphirom 2005: 188)

It is common for a sentence to have speech-level marking particles, question particles and pragmatic particles at the same time. In this circumstance, the particles usually appear in the sequence of pragmatic particle, question particle, and finally speech-level marking particle.

The linking particle /kô/ or /kôʔ/ has multiple functions. For example, it is used as a conjunction, equivalent to the English ‘so’, as in Example 1.11.

Example 1.11

yaŋ sàŋ iik | kô læy loŋ maa
yet shake one.more | LP LINK descend come/DIR
‘There was another tremor, so I came downstairs.’

Thai has one pre-verbal modal particle: /cà/. /cà/ is described as a modal particle, rather than a modal auxiliary, because it has a more general meaning, and it is often used alongside modal auxiliaries without making a big difference to the overall modality. The particle /cà/ indicates that the proposition the sentence expresses is challengeable, that is, that the speaker assumes that a hearer may not accept the proposition as a fact. Given the implication of challengeability, this particle is also called a challengeability marker. The

particle has multiple functions. It is, for example, used to indicate the challengeability of a future event, as in Example 1.12.

Example 1.12

léw mây lúu cà tham ηay dii
 LINK NEG know CM do how good
 ‘And I don’t know what to do.’

1.3.3.3 Word order

Thai sentence structure generally conforms to subject + verb + object word order (Campbell and King 2013: 1667). Whether the verb is transitive, intransitive, or ditransitive, it usually follows the subject of a sentence. There are at least two intransitive verbs, however, that normally precede their subject: /mii/ ‘(something) exists’ and /kəət/ ‘(something) occurs’.

Sometimes either the subject or object or even both are omitted (Smyth 2002: 117). That is, subject and object pronouns are not compulsory, as in Example 1.13.

Example 1.13

tôη rîp pay sáa hây
 must hurry go buy give
 ‘I must rush off and buy some for her.’

However, topicalisation, where a word or phrase other than a subject is placed at the beginning of a sentence, is a common feature in Thai (Smyth 2002: 117), as in Example 1.14.

Example 1.14

sâa kàw ca aw pay bəricàak phrûη ní
 clothes old will take go donate tomorrow
 ‘I’ll give away the old clothes tomorrow.’

In the written language, however, /sùan/, /sǎmràp/, or /rǎaŋ/ is often used to introduce the topic, which is often followed by /nán/, if it is long. The verb is then introduced by /kǎ/.

Another common pattern in Thai is to have a pronoun immediately after a subject noun, which it recapitulates (Smyth 2002: 116), especially in the spoken language, as in Example 1.15.

Example 1.15

rót	man	tit
cars	they	stuck
‘The traffic is jammed.’		

1.4 Overview of the thesis

This thesis will be divided into seven chapters. This section is a brief overview of what each chapter will cover.

Chapter 2 covers the theoretical background relevant to the research questions outlined in section 1.2. I will first provide some background on neo-Firthian corpus linguistics, wherein the concept of semantic prosody is situated (in section 2.2). I will then discuss the approaches to semantic prosody established in the literature (in section 2.3). This will include a detailed discussion of Louw, Stubbs, Partington and Sinclair’s approaches. This will be followed by a brief overview of Hoey’s concept of lexical priming (in section 2.4), as this is a concept relevant to the discussion of semantic prosody. The foregoing reviews will then lead into a discussion of the similarities and differences among the contrasting approaches to semantic prosody (in section 2.5).

Finally, in section 2.6 and section 2.7, I will review a range of earlier studies of semantic prosody both in English and other languages (including contrastive studies).

Chapter 3 discusses the methodology. I will first provide a brief introduction to the data to be used in the study, which is the Thai National Corpus and the British National Corpus (in section 3.2 and section 3.3). I will then discuss the methodology to be applied in each of the two primary approaches to semantic prosody that have been identified in the literature review. The first method will be based on Louw, Stubbs and Partington's approaches (section 3.4.1). Within this approach, semantic prosody will be identified based on the word's collocates, and it will be restricted to the positive vs. negative opposition. The second method will be based on Sinclair's approach (section 3.4.2). Within this approach, semantic prosody will be identified from extended co-text around the word, and it can be any pragmatic function or meaning rather than being restricted to the positive vs. negative opposition.

Chapter 4, Chapter 5, and Chapter 6 present the results of the analyses directed at answering research question 1, research question 2, and research question 3, respectively. In each of these chapters, I will present the approach(es) adopted in the analysis, and also the rationale for the selection of the given approach(es). I will also present the words whose semantic prosody will be analysed, as well as the criteria for the selection of those words. The results of the analysis will be discussed in detail.

In Chapter 4, I will examine the advantages and disadvantages of the two major approaches to the study of semantic prosody in Thai. This will be accomplished by investigating three Thai verbs using both approaches (in section 4.4.1 and section 4.4.2).

The results of the analysis will lead to a detailed discussion of the differences between the two approaches and their advantages and disadvantages (in section 4.5).

In Chapter 5, I will investigate variation in semantic prosodies across genres. This will involve the identification of an extended unit of meaning around 19 verbs in four different genres (in section 5.5). The variation in the use of the extended units of meaning identified will lead to a discussion of variation in semantic prosodies across genres (in section 5.6.5).

In Chapter 6, I will examine the extent to which the semantic prosodies of translation-equivalents in Thai and English are similar or different. This will involve a collocate analysis of 48 Thai nouns and their English translation-equivalents (in section 6.5). Finally, the similarities and differences in semantic prosodies of the pairs under study will be discussed (in section 6.6).

Chapter 7 concludes the thesis. I will first provide a summary of the findings of the study and lay out the answers to my research questions (in section 7.2). I will then go on to discuss the limitations of the study (in section 7.3). This will then be followed by a statement of the novel contribution made by the study (in section 7.4) and my suggestions for some directions for possible future research (in section 7.5).

Chapter 2 – Literature review

2.1 Introduction

The purpose of this chapter is to review the theoretical background relevant to the research questions outlined in section 1.2. Section 2.2 gives some background about neo-Firthian corpus linguistics, in which the concept of semantic prosody is situated. Section 2.3 discusses the various approaches to semantic prosody existing in the literature. In section 2.4, a brief overview of Hoey's lexical priming theory is given, as this concept is relevant to the discussion of semantic prosody. In section 2.5, similarities and differences among the approaches to semantic prosody are discussed in detail. Finally, sections 2.6 and 2.7 review studies of semantic prosody in English and in other languages respectively.

2.2 Neo-Firthian corpus linguistics

2.2.1 Introduction to neo-Firthian corpus linguistics

Neo-Firthian corpus linguistics is a branch of corpus linguistics associated with scholars operating within, as its label suggests, J. R. Firth's approach to the study of language (McEnery and Hardie 2012: 122). One of the first to apply a corpus linguistic methodology to Firth's approach was John Sinclair, who was Professor of Modern English Language at the University of Birmingham. Other neo-Firthian scholars are mostly associated with the University of Birmingham; these include Susan Hunston, Bill

Louw, Michael Hoey, Michael Stubbs, Wolfgang Teubert, and Elena Tognini-Bonelli (McEnery and Hardie 2012: 122).

Neo-Firthian corpus linguistics is closely associated with the corpus-driven approach (McEnery and Hardie 2012: 6), as opposed to the corpus-based approach. The corpus-driven approach views corpus linguistics as a theory, and thus claims that “the corpus itself embodies its own theory of language” (McEnery and Hardie 2012: 6). To put it another way, corpus-driven linguists approach the corpus with no pre-existing theory in mind, aiming to postulate a new theory of language based on the corpus data (McEnery *et al.* 2006: 10). For this reason, they strongly object to corpus annotation, which is the process of adding interpretative linguistic information to the corpus, such as part-of-speech tagging (McEnery *et al.* 2006: 29). Moreover, corpus-driven linguists do not make an attempt to attain corpus balance and representativeness when designing and building corpora; rather, they believe that once a corpus grows large enough, it will achieve cumulative representativeness (McEnery *et al.* 2006: 8). This might be the reason why they also argue in favour of very large corpora. For example, the Bank of English (BoE), created by scholars at Birmingham in collaboration with Collins publishers, contains 650 million words.

Semantic prosody is a concept extensively discussed by neo-Firthian corpus linguists. It was introduced to the public by Bill Louw (see section 2.3.1.1), and is notably regarded as a compulsory element of Sinclair’s model of extended units of meaning (see section 2.3.2.1). The notion is also debated by many neo-Firthian corpus linguists, including Hunston, Hoey, Stubbs, and Tognini-Bonelli. In the next section, I will discuss Sinclair’s notion of the idiom principle, which leads to his proposal of

extended units of meaning and his conception of collocation, two ideas that are invariably relevant in any discussion of semantic prosody.

2.2.2 Open-choice principle vs. idiom principle

Sinclair (1991: 109) advances two models for how meaning arises from a language text: the open-choice principle and the idiom principle. Under the open-choice principle, which Sinclair also refers to as ‘slot-and-filler’ model, texts are seen as “a series of slots which have to be filled from a lexicon which satisfies local restraints” (Sinclair 1991: 109). Under this model, each slot can be filled by a large number of possible choices, and “the only restraint is grammaticalness” (Sinclair 1991: 109). Most traditional models of grammar are the kind of grammar Sinclair calls a ‘slot-and-filler’ grammar. One example is Chomsky’s phrase structure grammar (Chametzky 2000: 3) where sentences are analysed into a hierarchy of constituents. That is, sentences are regarded as “objects analysable into ever-smaller constituent parts until the ultimate constituents had been reached” (Horrocks 1987: 31). These smallest parts together form the various kinds of phrases that constitute a sentence. The formation of phrases, with regard to their constituents and their internal order, is in turn governed by a set of phrase structure rules (Horrocks 1987: 31-32). In one early version of this theory, each sentence consists of a noun phrase and a verb phrase. This noun phrase and this verb phrase can each in turn be further analysed into constituents and sub-constituents, creating a branching hierarchy. This analysis is often presented in the form of a tree diagram (Horrocks 1987: 32). The tree diagram in Figure 2.1 shows how the sentence *Those boys like cars* can be analysed in this theory of grammar.

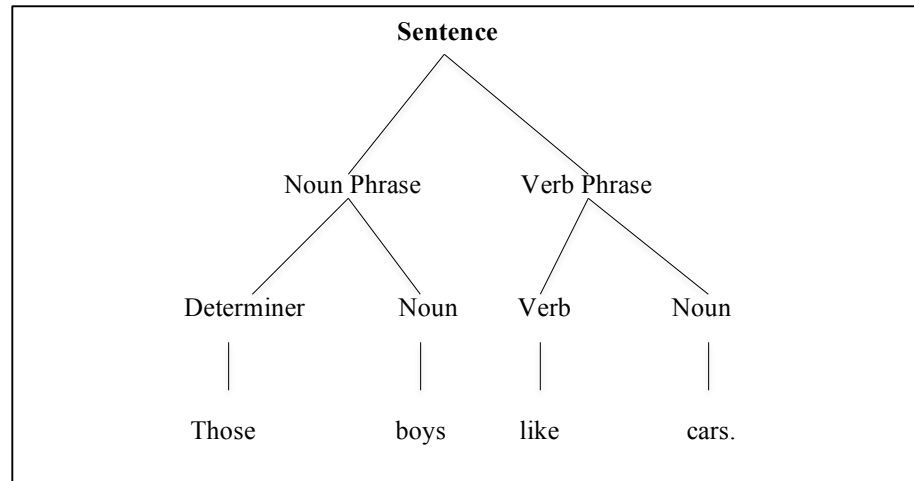


Figure 2.1 An example of a phrase structure tree diagram

Under this model, each of the category labels at the bottom of the sentence tree can be seen as a slot that can be filled by any word, which is independently selected as long as it satisfies local grammatical constraints, as discussed earlier. Therefore, having the words *those*, *boys*, *like*, and *cars*, as well as *Those boys like cars*, we can compose the sentence *Those cars like boys*, each part of which *also* satisfies the relevant local grammatical restraints. In the first phrase, for instance, *those* is a determiner and *cars* is a noun. The combination of these words in the determiner-noun order adheres to the phrase structure rules, forming a noun phrase. This noun phrase in turn functions as a subject of the verb phrase *like cars*, which also follows the syntactic rules. Together the two phrases form a well-formed or grammatical sentence.

From the analysis of the constituents of these two example sentences, it also follows that the meaning of a sentence is derived from the combination of the meaning of each individual word, according to the principle of compositionality (Löbner 2002: 13-15). That is, a sentence's meaning is determined by the combination of the lexical meaning, or the meaning stored in the lexicon, of each word, guided by the grammatical

structure of the sentence. Under this principle, the sentence *Those cars like boys*, just like the sentence *Those boys like cars*, is not only well-formed but able to convey a compositional meaning, even though the sentence *Those cars like boys* does not seem to make much sense in practice.

However, Sinclair (1991: 110) argues that operating by this open-choice principle alone cannot produce a normal text, as words in fact do not occur randomly subject only to formal grammatical constraints. As Sinclair (1991: 110) puts it, “the nature of the world around us is reflected in the organization of language and contributes to the unrandomness.” This is where the idiom principle comes into play. By the *idiom principle*, Sinclair means:

A language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments.

(Sinclair 1991: 110)

Under this principle, certain words are likely to go together and “make meaning by their combinations” (Sinclair 2004: 29); that is, there exist prefabricated phrases in the lexicon, each of which is considered one unit of meaning. For example, *of course*, though analysable into two separate words, in fact constitutes a single lexical choice and makes one unit of meaning (Sinclair 1991:110).

Sinclair (1991: 112) further argues that idiomaticity is not a minor feature, as it is usually regarded in most current linguistics. Idiomaticity is traditionally associated with semantically non-compositional multi-word units such as idioms, where “the individual words have ceased to have independent meanings” (Saeed 1997: 60). In Sinclair’s view, by contrast, the realm of idiomaticity is extended to include, alongside idioms, semantically compositional multi-word expressions and combinations, such as fixed n-

grams or clusters and more variable collocations, the latter of which will be discussed in detail in the next section.

Sinclair goes on to suggest that the idiom principle is the primary model that accounts for how meaning arises in normal texts, as “most of the text will be interpretable by this principle” (1991: 114). Nevertheless, the open-choice principle cannot be discarded. As Sinclair (1991: 114) suggests, “Whenever there is good reason, the interpretive process switches to the open-choice principle.” There also exist texts that are probably composed in a fashion that “makes greater than normal use of the open-choice principle” (Sinclair 1991: 114), such as legal statements. Hence, even though the idiom principle dominates, the open-choice principle still has a role to play.

2.2.3 Collocation

2.2.3.1 Introduction to collocation

Collocation is an old concept (McEnery and Hardie 2012: 122) and has been variously defined (Bahns 1993: 57; Chi *et al.* 1994: 157; Nesselhauf 2005: 11). However, one aspect that most scholars seem to agree upon is that collocation refers to some kind of syntagmatic relation (Nesselhauf 2005: 11). Below is a brief overview of some approaches to collocation in the literature.

One approach to collocation views it as a type of word combination. This approach, which Nesselhauf dubs the phraseological approach, is usually adopted by scholars working in the areas of lexicography and/or pedagogy (Nesselhauf 2005: 12). Within this approach, collocation is usually delimited from other types of word

combination on the basis of transparency (whether the elements of the combination and the combination itself have a literal or a non-literal meaning) and commutability (whether substitution of the elements is possible and to what degree). For example, Cowie (1981 cited in Nesselhauf 2005: 14) defines collocation (which is also referred to as *restricted collocation*) by distinguishing it from “free combination” on the one hand and from “idiom” (which he further divides into figurative idiom and pure idiom) on the other. This approach also requires the components of the combination to be syntactically related; thus, analyses are usually conducted on pre-defined syntactic structures, such as adjective + noun, noun + noun, and adverb + adjective (Nesselhauf 2005: 17).

Collocation is also considered from the perspective of patterns of co-occurrence. Within this approach, collocation is studied under different rubrics, such as n-grams, clusters, or lexical bundles (McEnery and Hardie 2012: 123). N-grams refer to sequences of n words, where n is variable; thus, we can have bi-grams (two words in length), tri-grams (three words), four-grams, and so on (Friginal and Hardy 2014: 41). Lexical bundles are an approach to the analysis of n-grams (Friginal and Hardy 2014: 41). The term “lexical bundles” is coined by Biber *et al.* (1999: 990) to refer to “sequences of word forms that commonly go together in natural discourse.” These sequences of two or more words must occur frequently in order to be considered lexical bundles, although the frequency cut-off may vary (Biber *et al.* 1999: 990). Generally, most lexical bundles do not have an idiomatic meaning and do not constitute a complete syntactic unit (Biber *et al.* 2004: 376-377). Some examples of common lexical bundles in conversation are *I don't know why* and *I thought that was* (Biber *et al.* 1999: 991). These bundles are

transparent in meaning, and they are structurally incomplete, in that they consist of a subject followed by a verb phrase and only the start of a complement clause.

These approaches are, however, not the approaches typically adopted by Sinclairian or neo-Firthian corpus linguists. In the next section, I will discuss the Sinclairian or neo-Firthian approach to collocation, which is our present concern of this study.

2.2.3.2 The Sinclairian or neo-Firthian approach to collocation

Collocation is one of the central ideas of neo-Firthian corpus linguistics. The Sinclairian or neo-Firthian approach to collocation originates from J.R. Firth. Firth proposes that a word has different levels of meaning, one of which is meaning by collocation (Carter and McCarthy 1988: 32). Firth defines this as follows:

Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words. One of the meanings of night is its collocability with dark, and of dark, of course, collocation with night.

(Firth 1957: 196)

Firth's notion of collocation is further developed by John Sinclair (1991: 170), who defines collocation as "the occurrence of two or more words within a short space of each other in a text". With this definition, it thus follows that collocation can be seen as patterns of co-occurrence. These patterns are, however, potentially loose; the co-occurring items, although they must be within a specified span, do not have to be adjacent or in any fixed order (McEnery and Hardie 2012: 123).

Sinclair *et al.* (2004: 10) define three technical terms to describe collocation: *node*, *collocate* and *span*. The node refers to the word under study. The collocate refers to

any one of the words standing close to the node within a span, which is a distance within which a node and collocates are said to form a collocation. The span is usually defined as a space of about four words to the right and four words to the left of the node (Nesselhauf 2005: 12). For example, in the sequence ...*There was a dog playing in the playground. It looked happy and...*, if we take *playground* as a node and adopt a span of four, *dog, playing, in, the, it, looked, happy, and* may all be considered collocates and form collocations with the node *playground*.

Sinclair *et al.* (2004: 10) go on to distinguish two types of collocation: significant collocation and casual collocation; Nesselhauf (2005: 12) points out that elsewhere Sinclair sometimes reserves the term “collocation” only for the former. Significant collocation is “regular collocation between two items, such that they co-occur more often than their respective frequencies, and the length of the text in which they appear, would predict” (Sinclair *et al.* 2004: 10). Statistical significance testing is needed to extract significant collocation. Casual collocation, by contrast, is non-significant; that is, the co-occurrence is simply due to chance. Therefore, in the above example, not all co-occurrences are likely to form significant collocation. The distinction Sinclair makes between significant collocation and casual collocation is important when any individual co-occurrence is considered to constitute collocation. However, when the term “collocation” is reserved exclusively for significant collocation, such a distinction is no longer needed, and casual collocation can be considered as simple co-occurrence, and not collocation at all. Both these terminologies are in wide use. I will return to the discussion of collocation vs. simple co-occurrence in section 2.5.2.1.

Even though neo-Firthian corpus linguists are broadly similar in their view of collocation as the co-occurrence of words in a certain span, differences emerge with regard to their operationalisations of collocation. Two techniques are primarily used by neo-Firthians to identify the collocates of a node: a statistical technique and a non-statistical technique (McEnery and Hardie 2012: 126).

With the statistical technique, which McEnery and Hardie (2012: 127) dub “collocation-via-significance”, lists of collocates are automatically generated via statistical significance testing of the difference between the frequency of a given word type in the vicinity of the node and its frequency elsewhere. This technique takes into account the role of frequency and “the importance of a more-than-random co-occurrence”; it therefore reduces the possibility that the collocates extracted are due to chance (McEnery and Hardie 2012: 125). However, a problem of subjectivity has been pointed out with this technique: the statistic chosen has a considerable influence on the resulting collocates (McEnery and Hardie 2012: 127).

By contrast, with the non-statistical technique, which is also dubbed “collocation-via-concordance”, statistical testing is not at all involved, although frequency data may be manually compiled. Instead, a concordance is extracted for the node; then, each concordance line is manually examined, and items that recur close to the node are identified as collocates with the assistance of the intuition of the linguist. This technique is adopted by most neo-Firthian corpus linguists, with the exception of Krishnamurthy and Hunston, who argue in favour of the use of statistical significance tests. In other schools of corpus linguistics, the use of statistical significance testing prevails (McEnery and Hardie 2012: 127).

2.3 Approaches to semantic prosody

The concept of semantic prosody is considered an important concept in corpus linguistics (Bednarek 2008: 119; Whitsitt 2005: 283), and it has attracted much interest from corpus linguists in the past 15 to 20 years (Stewart 2010: 6). Given that it is a relatively new concept, there is no consensus on its definition (Zhang 2009: 2). In fact, the term has been used and defined differently by scholars such as Louw (1993), Sinclair (2004), Stubbs (1995; 2001), and Partington (1998; 2004; 2014).

2.3.1 Louw's approach to semantic prosody

2.3.1.1 Louw's introduction of semantic prosody

Louw was the first person to introduce the term semantic prosody to the public (Whitsitt 2005: 283). However, Louw (1993: 158) expresses gratitude to Sinclair for having provided him with both the term and the concept in their personal communication. In Sinclair's earlier work, he describes a phenomenon that seems to be semantic prosody though he does not explicitly use that term. For example, his observations on the phrasal verb *set in* (Sinclair 1987: 155-156) and the verb *happen* (Sinclair 1991: 112) demonstrate that both are habitually associated with unpleasantness. Regarding *set in*, Sinclair notes:

The most striking feature of this phrasal verb is the nature of the subjects. In general they refer to unpleasant states of affairs. Only three refer to the weather; a few are neutral, such as *reaction* and *trend*. The main vocabulary is *rot* (3), *decay*, *malaise*, *despair*, *ill-will*, *decadence*, *impoverishment*, *infection*, ...

(Sinclair 1987: 155-156)

Of *happen*, Sinclair says:

Many uses of words and phrases show a tendency to co-occur in a certain semantic environment. For example, the verb *happen* is associated with unpleasant things – accidents and the like.
(Sinclair 1991: 112)

The phenomenon of *set in* and *happen* being habitually associated with unpleasant states of affairs is, according to Louw (1993: 158), referred to by Sinclair as *semantic prosody*. Specifically, *set in* and *happen* are found to have a negative semantic prosody because they habitually co-occur with items that have negative meanings. Louw (1993: 158) goes on to say that Sinclair uses the term “prosody” in the same sense as Firth when the latter uses it to explain phonological coloring. To explain this, Louw uses the word *Amen* as an example. The nasal consonants /n/ and /m/ in the word *Amen* are capable of colouring the non-nasal vowels with a nasal quality by virtue of their adjacency. Following Firth’s concept of phonological coloring, Louw thus argues:

The habitual collocates of the form *set in* are capable of coloring it, so it can no longer be seen in isolation from its semantic prosody, which is established through the semantic consistency of its subjects.
(Louw 1993: 159)

It can thus be argued that Louw views semantic prosody as the result of *meaning transfer* (to use Whitsitt’s terminology). That is, the meaning of the collocates of the phrasal verb *set in* is transferred to *set in*. This transfer results in *set in* consistently carrying the meaning transferred to it from its “unpleasant” collocates. That is, it results in *set in* consistently carrying a bad semantic prosody. This concept of meaning transfer is also arguably implicit in the definition Louw (1993: 159) gives to semantic prosody as “a consistent aura of meaning with which a form is imbued by its collocates.” This definition may be read as suggesting that once a form is imbued by its collocates’

meaning, it always carries that particular meaning as its semantic prosody. This is an issue first highlighted by Whitsitt as a criticism of Louw. Louw (1993: 164) goes on to argue that semantic prosody is a result of “a long period of refinement through historical change,” again suggesting that this meaning transfer is a historical process.

Even though, as previously mentioned, Louw expresses gratitude to Sinclair for having provided him with both the term and the concept of semantic prosody, Louw’s concept of semantic prosody is not exactly the same as Sinclair’s. (Sinclair’s account of semantic prosody will be discussed in section 2.3.2.1.) In Sinclair’s account of *set in*, for example, he just points out that the phrasal verb is habitually associated with unpleasantness. He does not refer to or imply a diachronic transfer of meaning. Thus, I argue that the idea of meaning transfer over time originates from Louw, not Sinclair. Whitsitt (2005: 284) makes a parallel argument, observing that for Sinclair, the idea of semantic transfer is “dramatically reduced”.

Louw (1993: 171) originally categorises semantic prosody into ‘good’ and ‘bad’ or ‘negative’ prosody. However, as subsequent work has standardised on the terms *positive* and *negative*, I will use these latter two terms. Investigating the expressions *utterly*, *days are*, *bent on*, *symptomatic of*, and *build up*, Louw finds that these items display the semantic prosodies summed up in Table 2.1.

Word/Expression	Semantic prosody
Utterly	Overwhelmingly negative prosody
Days are	Negative prosody
Bent on	Negative prosody
Symptomatic of	Negative prosody
Build up	Positive/negative prosody

Table 2.1 Words/Expressions studied by Louw and their prosodies

Specifically, Louw's investigation of a concordance for *utterly* drawn from the Cobuild corpus shows that the right-collocates of *utterly* are mostly unpleasant, resulting in the word carrying an "overwhelmingly 'bad' prosody" (Louw 1993: 160). Some of its unpleasant collocates are *stupid*, *ridiculous*, and *insensible*. Similarly, his observation of the phrase *days are* in the same corpus shows that it is usually followed by negative words like *gone*, *over*, and *past*. The concordances for *bent on* and *symptomatic of*, similarly drawn from the Cobuild corpus, also show that both are frequently followed by negative things, e.g. *destroying*, *mischief*, and *revenge* for *bent on*; as Louw (1993: 166) puts it, "the pursuits that people are *bent on* are almost always negative or unpleasant in some way." *Build up* is, however, a special case; it displays either a positive or a negative prosody depending on whether it is used transitively or intransitively (Louw 1993: 171). That is, it is observed to carry a positive semantic prosody when it is used transitively with a human subject (as in *people build up better understanding*), but it is found to display a negative prosody when it is used intransitively (as in *toxins build up*).

It can be seen from Table 2.1 that all of the expressions that Louw investigates display a negative semantic prosody. Louw (1993: 171) in fact suggests that there seem to be "more 'bad' prosodies than 'good' ones." He goes on to argue that semantic prosodies prove useful for stylistics and the analysis of suasive writing, and that they are "inaccessible to human intuition about language and they cannot be retrieved reliably through introspection" (Louw 1993: 157).

2.3.1.2 Louw's binary claim about semantic prosody

Louw (1993: 169) makes one further important claim about semantic prosody. He argues that semantic prosody serves as a tool for a writer or speaker to achieve an ironic effect. That is, when a writer or speaker produces a word within a combination that goes against the established semantic prosody of that word, they intend their message to be interpreted ironically. Louw (1993: 164) uses *bent on* as an example to illustrate this point. He considers the phrase “bent on self-improvement” in David Lodge’s novel *Small World* and argues that it reveals Lodge’s ironic intention. Louw posits that by collocating *bent on*, which is established to have a negative prosody and is thus expected to collocate with negative words, with *self-improvement*, an apparently positive word, Lodge deliberately shows his ironic intention. That is, Lodge does not think that self-improvement is a good thing in this context. However, Louw (1993: 164) also posits that semantic prosody can be employed to instantiate an ironic effect only when a word’s prosody and collocational behaviour are strongly established. As he puts it, “there must be a sufficiently consistent background of expected collocation against which the instantiation of irony becomes possible” (Louw 1993: 157).

Louw (1993: 169) further argues that there may occur a situation where a writer or speaker’s usage runs contrary to an established semantic prosody, but there is apparently no sign of an intention to be ironic. In this circumstance, he posits that the prosodic clash can be interpreted as resulting from the writer or speaker’s attempt to hide their real attitude. Louw (1993: 169) looks at the use of *symptomatic of* in a live interview with an official called Richard Francis on Zimbabwe Television in the following passage as an example to illustrate this point.

Francis: Well, it's very wide. I mean, it's *symptomatic of* the University of Zimbabwe which has such a high reputation that there are fifteen links between departments in the university here and equivalent departments in all sorts of situations, universities, polytechnics in Britain.

Louw (1993: 169)

Louw (1993: 169-170) argues that here Francis, who was the Director General of the British Council, does not actually have a high opinion of the University of Zimbabwe despite his apparently praising the University for having *a high reputation*; and he argues that this is exposed by the prosodic clash. That is, the use of *symptomatic of*, which is established to carry a negative prosody, with the obviously positive phrase *a high reputation* reveals the speaker's true opinion of the University: he does not believe that the University has high standards and a high reputation. In fact, he feels that the University desperately needs help from Britain, as seen in his mention of the 15 links between the University and departments in Britain. But as he does not wish to make a negative comment on the University on Zimbabwe Television, he instead skillfully suggests that because of the University's high standards, it has connections with the departments in Britain. In spite of this, his real opinion of the University is unavoidably revealed by his choice of *symptomatic of*. Louw (1993: 170) argues that there is no sign of irony in the context, ruling out the possibility that the speaker intends to be ironic. He likewise argues that there is no possibility of the speaker having made a slip of the tongue, considering his high linguistic proficiency. Thus, Louw (1993: 170) concludes that here the prosodic clash is the result of the speaker's attempt to conceal his true feelings.

We have, then, seen that Louw makes two important arguments about semantic prosody. First, semantic prosody is a result of a diachronic phenomenon (labelled by

others as a “transfer of meaning”), that is, the transfer of the collocates’ meaning to the terms next to them over time. This results in these terms consistently carrying the meaning transferred to them as their semantic prosody, which is as either positive or negative. Second, if a word is produced in a combination where it runs contrary to its typical semantic prosody, Louw argues that this prosodic clash can only indicate the speaker or writer’s ironic intention or insincerity. These two arguments have, however, been criticised by other scholars, particularly by Whitsitt and Hunston.

2.3.1.3 Whitsitt’s criticism of Louw’s characterisation of semantic prosody

Whitsitt (2005: 284) finds Louw’s characterisation of semantic prosody as a diachronic transfer of meaning from the collocates to the terms standing near them to be problematic. He reasons that this idea of meaning transfer assumes that some words are full of meaning, whereas others are empty of meaning. Particularly, Whitsitt (2005: 290) argues that such an idea implies that “collocates are being thought of as content without form, while the term to be imbued is form without content.” That is, Whitsitt believes that Louw suggests that the terms imbued with meaning by their collocates were empty of meaning to begin with. However, Louw does not explain why some words – specifically collocates – are full of content or meaning, while others – specifically the terms standing next to the collocates – are empty of meaning. In addition, as Sinclair (1996: 115 cited in Whitsitt 2005: 291) points out, assuming such a claim is valid, an explanation on why there must be a “spillover of meaning” from the collocates to the terms next to them is still needed.

Whitsitt (2005: 287-288) also attacks Louw's use of synchronic corpora in his investigations of semantic prosody. He reasons that, given that a transfer of meaning over time is central to Louw's concept of semantic prosody, it is invalid to make claims about a diachronic transfer of meaning based on observations derived from synchronic corpora. Thus, for Whitsitt, Louw's claims regarding semantic prosody are suspect and should not even merit any attention in the first place, as Whitsitt (2005: 287) puts it, "one could make the claim that there are valid grounds for dismissing semantic prosody from the outset as a concept worthy of study." McEnery and Hardie (2012: 139) concur with Whitsitt on this point, asserting that he "accurately points out that concordances from a corpus of contemporary English cannot, by their very nature, provide evidence for such a process of change over time."

In addition to his criticisms of the underlying idea of semantic prosody, Whitsitt criticises the analogies Louw uses in defining the concept. In particular, Whitsitt (2005: 291) attacks the analogy Louw draws between Firth's concept of phonological coloring and semantic colouring, and his subsequent borrowing of the term *prosody* to describe the latter phenomenon. That is, he argues that these phenomena are not similar in the relevant sense, because although nasal consonants are capable of colouring adjacent vowels with a nasal quality, for instance, in the word *amen*, "the vowels do not get permanently colored with a nasal sound." Unlike what Louw argues to be the case in semantic prosody, nasalisation does not necessarily mean that the vowels, once nasalised, can thereafter only co-occur with nasal sounding consonants. On these grounds, Whitsitt argues that Louw's claim about *set in* (that once it gets imbued with a negative meaning, it invariably carries that meaning with it) is invalid. Rather, Whitsitt asserts that such a form, once

imbued with a negative meaning, does not necessarily always carry the unpleasant meaning. It is not *always* likely for such a form to occur only with unpleasant words or to colour other words with its negative prosody.

I would argue that Whitsitt's criticism of Louw in this regard is unsound. Louw (1993: 158) clearly states that it is Sinclair who coins the term *semantic prosody*, borrowing the term *prosody* from Firth's concept of phonological colouring. Thus, it would be unfair for Louw to be criticised for this. The term *prosody* is just borrowed to label the phenomenon; it is not the underlying idea of the phenomenon. What operates in phonological colouring of *amen* may or may not operate in semantic colouring of *set in*. Thus, to invalidate the concept of semantic prosody on these grounds would be unreasonable. As McEnery and Hardie (2012: 139) argue, Whitsitt's attacks on the analogies Louw uses to present the concept of semantic prosody are "ill-founded", and the concept "must stand or fall on its own merits rather than those of the analogies used to present it."

2.3.1.4 Hunston's criticism of Louw's binary claim about semantic prosody

Louw's argument regarding prosodic clashes is also questioned. McEnery and Hardie (2012: 140) argue that Louw makes a binary claim about semantic prosody. That is, when a speaker or writer goes against the established semantic prosody of a word, using it in an atypical environment, they are either being ironic or being insincere. This claim is problematic, as Hunston (2007: 261) argues. Specifically, she argues that a prosodic clash may indicate neither irony nor insincerity, but may rather simply be an atypical usage. Hunston (2007: 261) uses the phrase *to the point of* to illustrate this. This

phrase is “used to link a less saturated evaluative item with a more saturated one” and is established to have a negative semantic prosody. Some examples found in a concordance are *abundant to the point of extravagance*, *suspicious to the point of paranoia* and *naïve to the point of idiocy*. Nevertheless, alongside these examples where the phrase is, as typically, used in negative contexts, Hunston finds one concordance line where it is used in a positive environment as follows:

At 23 and with just one exhibition behind her, Brisbane painter Hazel Dooney is fresh **to the point of** invigoration. Her bold and colorful paintings combine elements of youth culture into pop art which is appealing not only for its pure aesthetics but also for its contemporary nature.

(Hunston 2007: 262)

Hunston (2007: 262) posits that in this example the writer is neither being ironic nor being insincere, pointing out that there is no evidence of either in the whole article. This refutes Louw’s binary claim that a prosodic clash must be indicative of either irony or insincerity on the part of the writer or speaker. Instead Hunston (2007: 262) argues that this prosodic clash – the co-occurrence of *to the point of* with *invigoration* – seems to be “simply an atypical use of the phrase.” This counter-example, as Hunston (2007: 263) argues, does not in any way invalidate the claim that *to the point of* typically co-occurs with a negative unit. It is rather the claim that the use of the phrase *to the point of* with a positive unit *always* indicates irony or insincerity that should be reconsidered.

In sum, although Louw’s definition of semantic prosody is the most frequently quoted in the literature to date (Zhang 2010: 190), it is also the definition that has been most severely criticised, as we have seen. Moreover, even though Louw claims that he follows Sinclair’s concept of semantic prosody, I find that his view is quite different from

that of Sinclair. In fact, in Sinclair's later work, his account of semantic prosody is completely different from Louw's, as will be seen in the next section.

2.3.2 Sinclair's approach to semantic prosody

2.3.2.1 Sinclair's introduction of semantic prosody

Sinclair's view of semantic prosody is closely associated with his proposal of a model of the extended unit of meaning (Steward 2010: 10). Sinclair (2004: 29-30) argues for the notion that "a linguistic item can be extended, at least for English, so that units of meaning are expected to be largely phrasal." This goes back to Sinclair's (2004: 29) proposal of the idiom principle, under which words are likely to go together and "make meaning by their combinations," as discussed in section 2.2.2.

To validate his claim that the units of meaning are largely phrasal rather than individual words, Sinclair (2004: 30-35) investigates the pattern of the idiom *naked eye* from a concordance drawn from the Bank of English. His analysis shows that the pattern on the left of the node is much more consistent than that on the right (Sinclair 2004: 31). This left-hand-side pattern can be summarised as follows (Sinclair 2004: 31-32):

see/visible	with/to	the	naked eye
N-3	N-2	N-1	node

Sinclair designates the position immediately to the left of the node as N-1, the position next to N-1 as N-2 and the one next to N-2 as N-3. The concordance lines show that *the* dominates N-1. Borrowing Firth's term, Sinclair (2004: 32) argues that this co-

occurrence between *naked eye* and the article *the* is an example of colligation or “the co-occurrence of grammatical choices.” Similarly, the prepositions *with* and *to* in N-2 are argued to be colligates, too. *See* and *visible* in N-3 are argued to form an instance of collocation, which is the co-occurrence of lexical choices. However, even though *see* and *visible* dominate N-3, there also exist concordance lines where N-3 is occupied by other words, all of which are either verbs or adjectives, such as *detect*, *spot*, *apparent*, *undetectable* and *evident*. Sinclair adds that this very restriction (to verb or adjective at N-3) is also a form of colligation. Based on all of these collocates, Sinclair further argues that *naked eye* has a semantic preference for *visibility*, pointing out as follows:

On this occasion colligation, being divided between the two, is not as important as another criterion, that of *semantic preference*. Whatever the word class, whatever the collocation, almost all of the instances with a preposition at N-2 have a word or phrase to do with visibility either at N-3 or nearby.

(Sinclair 2004: 32)

Semantic preference is “the restriction of regular co-occurrence to items which share a semantic feature” (Sinclair 1998: 16). Thus, in this example, the idiom is argued to have a semantic preference for visibility because *see* and *visible* and almost all of the other words that occupy N-3 or a nearby position belong to the same semantic field of visibility.

Closely examining the left-hand-side context of N-3, Sinclair (2004: 33) goes on to argue that the idiom *naked eye* has one more important element – a semantic prosody.

Sinclair here defines semantic prosody as follows:

A semantic prosody (Louw 1993) is attitudinal, and on the pragmatic side of the semantics/pragmatics continuum. It is thus capable of a wide range of realization, because in pragmatic expressions the normal semantic values of the words are not necessarily relevant. But once noticing among the variety of the

expression, it is immediately clear that the semantic prosody has a leading role to play in the integration of an item with its surroundings. It expresses something close to the ‘function’ of the item – it shows how the rest of the item is to be interpreted functionally. Without it, the string of words just ‘means’ – it is not put to use in a viable communication.

(Sinclair 2004: 34)

Even though Sinclair cites Louw, I would argue that his account of semantic prosody is in fact different from Louw’s. Sinclair discusses the function of semantic prosody, arguing that semantic prosody is “attitudinal, and on the pragmatic side of the semantics/pragmatics continuum.” Louw, by contrast, does not discuss the attitudinal and pragmatic function of semantic prosody. What he seems to focus on, as has been seen, is a transfer of positive or negative meaning and the use of prosodic clash to indicate irony or insincerity, none of which is discussed or implied by Sinclair. That said, Louw’s discussion of the transfer of positive or negative meaning does imply evaluation, which in turn implies attitudinal meaning – so this aspect is, in fact, present. I will discuss the differences between Louw and Sinclair’s accounts of semantic prosody in detail later.

Going back to the *naked eye* example, Sinclair (2004: 33) argues that *naked eye* has a semantic prosody of *difficulty*, pointing out that this prosody may be expressed by a word like *small*, *faint*, *weak* and *difficult* alongside a verb like *see*. For example, in the concordance line *too faint to be seen with the naked eye*, Sinclair (2004: 33-34) argues that some kind of difficulty is expressed jointly by the words *faint* and *seen*.

Sinclair’s notion of semantic prosody is further explained by McEnery and Hardie (2012: 138) who assert that “semantic prosody links the node to some expression of attitude or evaluation which may not be a single word, but may be given in wider context.” For example, they point out that in the unit *too faint to be seen with the naked*

eye the expression of difficulty, or to be more specific, difficulty with visibility, is “not evident from single words in the context of the node, but rather by a pragmatic interpretation.” In other words, unlike the semantic preference for *visibility*, which can be identified from the presence of the verb *seen*, there are not any single words in the unit that can be “identified as belonging to the semantic field of *difficulty*” (McEnery and Hardie, 2012: 138). Rather, the semantic prosody of *difficulty with visibility* is expressed pragmatically across the whole unit and has to be interpreted by a reader.

Having identified the semantic prosody, Sinclair (2004: 34) concludes that *naked eye* and its pattern on the left-hand-side context exemplify a “compound lexical item” or a “model of a lexical item consisting of several words.” This reaffirms his hypothesis, mentioned earlier, that a unit of meaning is largely phrasal. Specifically, Sinclair (1998: 14) argues that under this model, a lexical item or a unit of meaning has five components: three are optional and two are obligatory. The optional components are collocation, colligation and semantic preference, and the compulsory components are semantic prosody and the core (i.e. the basic word or words of the unit). Regarding the three optional components, Sinclair also (1998: 16) argues that they are “related to each other in increasing abstraction,” starting from collocation, colligation to semantic preference. Collocation is directly recognisable in the physical text. To identify colligation, however, one has to examine each word under study and assign each of them a grammatical class to see whether there is a predominant grammatical class. Finally, analysing semantic preference involves noticing the meaning of each word to see whether they belong to the same semantic set. In regard to the obligatory components, Sinclair (1998: 15) asserts that the core is “invariable, and constitutes the evidence of the occurrence of the item as a

whole,” and semantic prosody is “the determiner of the meaning of the whole.” He goes on to argue that it is in fact the semantic prosody of a lexical item that motivates a writer or speaker to produce that item in the first place (Sinclair 1998: 20).

In the *naked eye* example, *see* and *visible* constitute collocation. *The, with, to* and the dominance of the parts of speech of verb and adjective at N-3 constitute colligation. Sinclair (2004: 33) argues that there is a correlation between the prepositional choice – *with* and *to* – and the lexical choice – particularly *see* and *visible* –, depending on the part of speech of the latter. That is, the verbs select *with*, whereas the adjectives select *to*. The predominance of lexical choices to do with visibility, usually expressed by the words *see* and *visible*, establishes the idiom’s semantic preference for visibility. This semantic preference, when considered along with the phrasing on the left of the lexical choice that expresses it, in turn expresses the semantic prosody of difficulty with visibility. Finally, *naked eye* is the core, the almost unchanging part of the lexical unit. All of these components form one lexical item or one single unit of meaning that consists of many words. This can be summarised as follows:

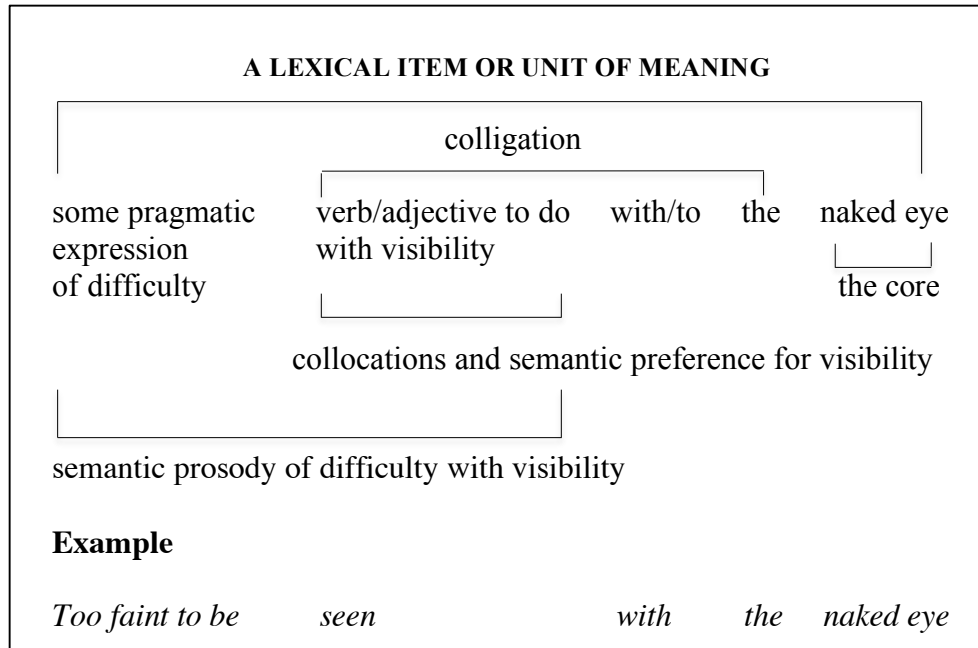


Figure 2.2 An extended unit of meaning whose core is *naked eye*

2.3.2.2 Hunston's support for Sinclair's view of semantic prosody

Sinclair's view on semantic prosody is embraced by Hunston. Hunston (2007: 251) points out that by semantic prosody, Sinclair means "the consistent discourse function of the unit formed by a series of co-occurrence." She also upholds Sinclair's notion that, as semantic prosody is an element of an extended unit of meaning, it resides in the whole unit or the sequence rather than in any particular word in the sequence. Hunston (2007: 257) supports this argument by referring to Sinclair's study of the word *budge*. She concludes from Sinclair's analysis that *budge* typically occurs in two sequences: 'inability (semantic preference) + negative (colligation) + *budge* (core) + (something)' (as in *Firemen...were unable to budge him*) and 'unwillingness (semantic preference) + negative (colligation) + *budge* (core)' (as in *The Prudential board refused to budge*). These whole sequences, as Sinclair (1998: 20) argues, express the sense of

“frustration (or a similar emotion) at the refusal or inability of some obstacle to move, despite pressure being applied.” This semantic prosody, summed up by Hunston (2007: 258) as “frustration in the face of difficulty,” belongs to the whole unit rather than to the word *budge* alone.

Hunston finds counter-examples to Sinclair’s generalisations of the use of *budge*. However, these counter-examples, as Hunston (2007: 257) argues, serve as a basis for her case in support of Sinclair’s proposal that semantic prosody is a property of an extended lexical unit. Hunston (2007: 257) considers the sequence *did not budge*, noting that although it usually indicates the semantic prosody of “frustration in the face of difficulty” as Sinclair argues, when the subject of the sequence is inanimate, it may also suggest “a positive achievement.” The example Hunston gives is as follows:

He [Otis] chose the New York Exposition to demonstrate his device, standing on an open lift high above the ground as his assistant cut the cord supporting it. The lift did not budge.

(Hunston 2007: 258)

Hunston argues that here the semantic prosody expressed by the sequence *the lift did not budge* is the audience’s surprise regarding the fact that the lift did not plunge to the ground, not the typical prosody of frustration.

Hunston (2007: 258) also notes the use of *budge* with the subjects *I* and *we*. Specifically, she argues that when the subject of the sequences is *I* as in “*The Prime Minister rejected resounding calls for the resignation of the government. ‘I will not budge,’ he said*” or *we* as in “*we will not budge*”, the semantic prosody expressed by the sequences is “determination in the face of the opposition rather than frustration” (Hunston 2007: 258).

Therefore, Hunston (2007: 258) asserts that although the word *budge* is typically observed to occur in fixed sequences, expressing the semantic prosody of “frustration in the face of difficulty,” this prosody is different when the elements of the sequence change; as she puts it, “If the phraseology changes, the semantic prosody is also different.” In particular, her further analysis of *budge* seems to suggest that the extended unit must also involve the subject. That is, as evident in the counter-examples, once the subject of the unit changes, the semantic prosody is different (which was not part of Sinclair’s original analysis). This thus in turn supports the argument that the discourse function or semantic prosody belongs to an extended unit of meaning rather than to a particular word.

In sum, Sinclair views semantic prosody as one of the elements of a compound lexical item or an extended unit of meaning, a unit that he argues to extend over several words. The semantic prosody is the attitudinal and pragmatic meaning of the unit. His view is supported by Hunston, who further argues that semantic prosody alters in accordance with changes in phraseology, reaffirming Sinclair’s concept that a semantic prosody belongs to an extended unit, not to a particular word. Sinclair’s proposal of the model of the extended unit of meaning and semantic prosody is also further developed by Stubbs, which I will discuss in the next section.

2.3.3 Stubbs' approach to semantic prosody

2.3.3.1 Stubbs' early study of semantic prosody

In an early work on semantic prosody, Stubbs (1995: 1) closely associates semantic prosody with collocation, which he defines as “a relationship of habitual co-occurrence between words.” Particularly, he suggests that a word's semantic prosody can be determined by its collocates and that quantitative methods should be adopted in identifying a word's collocates.

Stubbs (1995:3) examines the lemma *cause* and finds out that it habitually occurs in unpleasant environments. He thus concludes that, due to its predominantly unpleasant collocates, the lemma has a negative semantic prosody. Specifically, Stubbs says:

CAUSE is near the stage where the word itself, out of context, has negative connotations. (AFFECT is already at this point.) The selection restrictions on CAUSE are not (yet?) categorical: it is not (yet?) ungrammatical to collocate CAUSE with explicitly positive words. But it is easy to see how an increase in frequency of use can tip the balance and change the system.

Stubbs (1995: 20)

He also investigates other lemmas that are semantically related to the lemma *cause*. These include the lemmas *affect*, *consequence*, *create*, *effect*, *happen*, and *reason*. From the investigation he concludes that not all these lemmas have a clear semantic prosody. For example, the lemmas *create* and *result* have a mixed prosody, while the lemma *reason* is “largely neutral” (Stubbs 1995: 16).

Thus, Stubbs' early account of semantic prosody is that it is a phenomenon closely associated with collocation. If a word's collocates are mostly positive, the word displays a positive semantic prosody. But if most of its collocates denote negative

meanings, the word then comes to have a negative semantic prosody. I would therefore argue that at this point Stubbs' account of semantic prosody is close to Louw's, even though there are also some differences. First, both associate semantic prosody with collocation. However, they identify collocates quite differently. Stubbs (1995: 5-14) employs quantitative measures such as I- and T-values (based on Mutual Information and the T-test of significance) to identify collocates when working with large corpora. In contrast, Louw studies each concordance line one by one, and manually identifies the collocate in each line. These two different methods for identifying collocates correspond to "collocation-via-significance" and "collocation-via-concordance" techniques discussed in section 2.2.3.2. The reason why Louw manually identifies collocates might be that his concordances are small. In fact, one could also argue that Stubbs does the same as Louw when he makes use of data drawn from small corpora. For example, in his study of the lemma *cause* both as a verb and a noun in LOB, Stubbs (1995: 4) manually identifies the lemma's collocates from the concordance lines. As he puts it "I studied CAUSE in a 1 million word corpus (LOB), and began to document the lexical set of its collocates by constructing a concordance of all occurrences". Second, both Stubbs and Louw describe semantic prosody as being positive and negative. However, whereas Stubbs argues that a lemma can also have a mixed or neutral prosody, Louw restricts semantic prosody to being only positive or negative, arguing that a prosodic clash can only imply irony or insincerity, an idea not discussed by Stubbs.

2.3.3.2 Stubbs' later study of semantic prosody

Stubbs' account of semantic prosody in his later work, however, seems to develop in another direction. Specifically, he later develops Sinclair's proposal of the extended lexical unit, listing four separate kinds of relation between possible constituents of an extended lexical unit as follows (Stubbs 2001: 87-88):

1) Collocation refers to a lexical relation in a form of "a node-collocate pair" (Stubbs 2001: 64). It is "directly observable and countable in texts" (Stubbs: 2009: 124).

2) Colligation refers to "the relation between a pair of grammatical categories or in a slightly wider sense, a pairing of lexis and grammar" (Stubbs 2001: 65). It requires grammatical class analyses and is thus not directly observable (Stubbs 2009: 125). For example, the word-form *cases* habitually co-occurs with a member of the grammatical category of quantifier, such as *some* or *many* in *in some cases* and *in many cases*.

3) Semantic preference is "the relation, not between individual words, but between a lemma or word-form and a set of semantically related words" (Stubbs 2001: 65). For instance, the word-form *large* regularly co-occurs with words belonging to a semantic set of "quantities and sizes" (Stubbs 2001: 65).

4) Discourse prosody "extends over more than one unit in a linear string" (Stubbs 2001: 65) and expresses speaker attitude.

Stubbs' position is very close to Sinclair's in that he argues both for an extended lexical unit and for four major components (alonside the core of the item), one of which is semantic prosody. However, Stubbs does not follow Sinclair in using the term *semantic*

prosody to refer to the pragmatic and discourse function of an extended lexical unit. Instead he changes the terminology to *discourse prosody*. The reasons he gives are as follows:

Several studies use the term ‘semantic prosodies’ (Louw 1993; Sinclair 1996). ‘Pragmatic prosodies’ might be a better term, since this would maintain a standard distinction between aspects of meaning which are independent of speakers (semantics) and aspects which concern speaker attitude (pragmatics). I will here prefer the the term ‘discourse prosodies’, both in order to maintain the relations to speakers and hearers, but also to emphasize their function in creating discourse coherence.

(Stubbs: 2001: 65-66)

Stubbs examines the discourse prosodies of a number of words. One of these is the lemma *undergo* (Stubbs 2001: 89). Examining 343 concordance lines, Stubbs identifies the first 20 most frequent collocates of the lemma. He then groups these collocates into different semantic groups such as medicine, training and testing, change, and involuntariness, each of which represents a semantic preference of the lemma. From his observation of the concordance, he comes to the conclusion that the lemma has a simple and typical pattern: “people involuntarily undergo serious and unpleasant events, such as medical procedures” (Stubbs 2001: 89). This pattern, as Sinclair (1996: 95 cited in Stubbs 2001: 91) argues, can be realised by a number of lexical variations. Stubbs further argues that the lemma displays two related discourse prosodies: “involuntary” and “unpleasant”. The prosody of “involuntary” is almost always expressed to the left of the lemma, mostly through the words *forced* and *required*. This prosody can also be implied by the context, as in *Police said he would undergo psychiatric examination*. The prosody “unpleasant” is regularly encoded to the right of the lemma. Similar to the prosody of “involuntary”, this prosody can also be either expressed by particular words or implied by

the surrounding text. Both prosodies express the attitudinal and pragmatic function of the extended lexical unit whose core is the lemma *undergo*. In fact, it is the attitudinal and pragmatic function expressed by the discourse prosodies that motivates a speaker or writer to make the utterance in the first place (Stubbs 2001: 92). Simply put, because a speaker or writer wishes to express involuntariness in doing something unpleasant, they choose the extended lexical unit whose core is the lemma *undergo*.

In conclusion, Stubbs' thinking on semantic prosody has undergone some change. His early account of semantic prosody primarily concerns analysis of individual collocates. In this account, a word's semantic prosody is restricted to being positive, negative, or mixed. His view on semantic prosody later changes. In later work, he develops his notion of semantic prosody within the framework of Sinclair's model of the extended lexical unit. Emphasising its discourse function, Stubbs suggests a change in terminology to "discourse prosody." The discourse prosody is no longer limited to being straightforwardly positive or negative, as seen from the *undergo* example. In a sense, then, Stubbs moves from an early position that is similar to Louw's to a later position that is similar to Sinclair's. The final point to make in this section is that although Stubbs' later position with reference to semantic prosody is very close to Sinclair's, I would argue that in fact there remains some difference in the two scholars' analysis of semantic prosody (see section 2.5.3).

2.3.4 Partington's approach to semantic prosody

2.3.4.1 Partington's early study of semantic prosody

In his early work, Partington's (1998: 66) concept of semantic prosody revolves around the notion of connotation. He contends that semantic prosody is an aspect of expressive connotation; it expresses the speaker or writer's positive or negative evaluation of what they describe (Partington 1998: 66). He goes on to assert that often a favourable or unfavourable connotation or semantic prosody is not contained in a single lexical item. It is rather expressed "by that item in association with others, with its collocates" (Partington 1998: 66). For example, the unfavourable connotation or semantic prosody of *commit* does not only reside in the item, but rather "over a unit consisting of the item and its collocates" such as *offences* and *serious crime* (Partington 1998: 67).

The notion of semantic prosody spreading across a sequence of lexical items is also emphasised in his discussion of the difference between semantic prosody and expressive connotation. Partington (2004: 132) notes that items such as *rightly* and *flabby* have an apparent in-built (un)favourable evaluation or expressive connotation. Therefore, the use of these items alone reveals the speaker or writer's positive or negative attitudes towards what they describe. By contrast, semantic prosody "spread[s] over a unit of language which potentially goes well beyond the single orthographic word and is much less evident to the naked eye" (Partington 2004: 131-132). That is, it resides in "the collocational patterns of items in a text" (Morley and Partington 2009: 150).

Partington's discussion of the lemma *set in* further demonstrates his concept of semantic prosody. Following Hoey's (2005) theory of lexical priming, Partington (2004:

132) argues that *set in* is “*primed* to co-occur with unpleasant events or processes.” In other words, he argues that *set in* is expected to co-occur with negative collocates, not with positive ones. (Hoey’s theory of lexical priming will be discussed in detail later.) On this basis, Partington (1998: 67-68; 2004: 132) then argues that since the lemma *set in* habitually co-occurs with negatively evaluated items, it comes to have an unfavourable prosody; therefore, the negative evaluation of the unit involving the lemma *set in* is expressed not only by the lemma itself but also by its environments or collocates.

We have thus seen that an item’s semantic prosody is dependent on whether its co-occurring items are positively or negatively evaluated. However, Partington points out that this is not a simple relationship.

Simply being primed to appear in the environment of collocates of a certain evaluative sense, good or bad, is not a sufficient condition for an item to acquire the same sense. If the relationship between the item and its collocate is one of opposition or detraction, then the combination does not acquire the evaluative sense of the collocates.

(Partington 2004: 154-155)

To illustrate this, Partington (2004: 155) notes that although items like *relief* and *ease* habitually co-occur with negatively evaluated items such as *debt* and *pain*, the overall meaning of, say, *relief of pain* is positive. Therefore, the item *relief* does not acquire the negative evaluative sense of its collocates, but rather remains a positively evaluated item. This example clearly shows that, as Morley and Partington (2009: 142) point out, the semantic prosody of an item is not “a simple reflection of the nature of its co-occurring items.” In fact, the overall evaluative meaning of the unit and the embedded evaluation of the item itself (i.e. the embedded positive evaluation of *relief*) also have to be taken into consideration. Morley and Partington (2009: 146) do in fact go on to argue

that they prefer to discuss cases like *relief of pain* where the item's own evaluative meaning is obvious in terms of *evaluative embedding* than *semantic prosody*. Such combinations can be represented using the notation of (*relief of [pain]*), where round brackets indicate “good” and square brackets indicate “bad”. In this combination, *pain* remains a bad thing, but the overall evaluation of the whole combination and the evaluation of *relief* itself are good.

This issue could in fact underline the difference between Partington and Sinclair's approaches to semantic prosody. (This will be discussed in detail in section 2.3.4.2.) Partington's argument that *relief* remains a positively evaluated item rather than being given a negative semantic prosody by the negative meaning of its collocates highlights the association he makes between semantic prosody and connotation. His assertion that if *relief* has a negative semantic prosody, it will contradict the established positive connotation of *relief*, clearly stems from the close association between these concepts in his view. On the other hand, if we instead follow Sinclair's approach, which does not emphasise a connection between semantic prosody and connotation, and analyse the extended unit of meaning whose core is *relief* based on Sinclair's idea of semantic prosody, we might reach the conclusion that the unit has the semantic prosody of some unpleasant thing coming to an end. Just as the semantic prosody of “difficulty with visibility” motivates the use of *naked eye*, then, the semantic prosody of “unpleasant thing ending” would motivate the use of the unit of meaning based on *relief*. In this view of semantic prosody, we see easily that an “unpleasant thing ending” prosody and positive evaluative connotation are not at all contradictory.

Partington notes three further properties of semantic prosody. First, semantic prosody can be of varying strength; that is, “some items have a stronger good or bad prosody than others” (Partington 2004: 153). For example, Partington (2004: 144) finds that of the four lemmas belonging to the semantic group of *happen*, *set in* has the strongest unfavourable prosody, whereas *happen*, *occur* and *take place* have increasingly weaker prosodies. By this he means that *set in* is most unlikely to occur in non-negative environments, and this unlikelihood is incrementally less strong with *happen*, *occur*, and *take place* respectively. On this basis, he further argues that prosodic meaning is “more obviously *probabilistic* than some other kinds of meaning” (Partington 2004: 153). That is, even though a lemma may be said to have, for example, a positive semantic prosody, this does not necessarily mean that that lemma can co-occur *only* with items denoting positive meaning. For instance, Partington (2004: 153) points out that although the lemma *bent on* is said by Louw to have a negative semantic prosody, it may also be used in “a neutral or positive environment,” and its occurrence in such environment does not inevitably imply irony or a breakdown of communication, contrary to Louw’s claim.

Second, there is an interrelationship between semantic prosody and syntactic structure (Partington 2004: 144). That is, as Partington points out, particular prosodies are typically realised by particular structures. To illustrate this, he refers to Stubbs’ (2001: 200) investigation of the lemma *accost*, in which he finds that the lemma’s highly unfavourable prosody is typically realised within the passive structure. Louw (1993: 171) likewise seems to discover the same phenomenon when he argues that the semantic prosody of *build up* depends on whether it is used transitively or intransitively (see section 2.3.1.1).

Third, “the quality and strength of the prosody of a good many items will differ from genre to genre or from domain to domain” (Partington 2004: 153). Partington refers to Stubbs’ analysis of the lemma *lavish* as an example to illustrate this. Stubbs (2001: 106) finds that in newspaper prose, the lemma displays an unfavourable semantic prosody implying “a disapproving connotation of ‘excessive’ wastefulness.” On the other hand, in the arts and entertainment fields, Partington (2004: 153) finds that the lemma is generally “a neutral-to-good word”.

2.3.4.2 Hunston’s criticism of Partington’s conceptualisation of semantic prosody and Morley and Partington’s response

We have seen that Partington has consistently maintained that semantic prosody spreads across a unit of meaning larger than a single word. But his analyses of particular cases of semantic prosody often appear to suggest that he may rather view semantic prosody as a property of an individual lexical item. This issue has been raised by Hunston (2007: 250-251), who argues that Partington’s discussion of semantic prosody focuses on a “consistent co-occurrence of (types of) linguistic items” and “prioritises semantic prosody as the property of a word.” These concepts, she argues, are questionable. Hunston’s criticism of Partington is, however, not unexpected, considering her argument in favour of Sinclair’s notion of semantic prosody as a component of an extended lexical unit (see section 2.3.2.1). Hunston in fact asserts that Sinclair and Partington’s notions of semantic prosody represent two schools of thought on the phenomenon, clearly stating her preference for Sinclair’s (Morley and Partington 2009: 144).

Hunston (2007: 252) rebuts Partington’s argument by pointing out that “a word which is used in a certain way in most contexts is not necessarily used in that way in all

contexts.” Investigating a concordance of the lemma *cause* drawn from the magazine *New Scientist*, she finds that there is “no particular attitude [that] can be discerned towards the entity brought into being” (Hunston 2007: 252). In other words, what is caused is neither desirable nor undesirable. Some examples of the caused entities are *Lyman alpha line, fluctuations, a burst of star formation, and the heat loss*, none of which indicates any particular evaluation. These findings, as Hunston suggests, contradict the previous notion that the lemma *cause* has a negative semantic prosody and at the same time invalidate what she understands Partington’s position to be, namely that semantic prosody is the property of a word. Her point is that if the semantic prosody – negative in this case – really is a property of the lemma *cause*, there should not exist circumstances where the lemma is used in contexts other than negative, as has been seen. These counter-examples lead Hunston to conclude that the lemma *cause* does not consistently co-occur with items denoting negative things. Rather, it tends to occur with negative items and thus indicate negative evaluation only when the entities that are caused clearly involve human or animate beings. These entities may include illnesses, emotions, or processes related to people. It is only in such cases, as Hunston argues, that evaluation is relevant.

These findings about the lemma *cause* lead Hunston (2007: 253, 256) to argue that the immediate co-text plays a major role in the attitudinal interpretation of a lexical item. Thus, when we observe that a lexical item regularly co-occurs with items with, say, a positive meaning, we should not immediately jump to the conclusion that that item has a positive semantic prosody. This is because, as evident in her observations, this co-occurrence is “unlikely to be totally uniform” (Hunston 2007: 256). We must therefore be careful about making statements that a word or a short multi-word unit has a semantic

prosody “unless more precise observations of co-text and phraseology are made” (Hunston 2005: 256).

Hunston therefore effectively implies that Partington’s treatment of semantic prosody is excessively simplistic, simply assigning it to a word or a phrase without considering co-text or surrounding phraseology. I would argue that such a claim would be too strong. In fact, as we have seen, in his study of the phenomenon, Partington also notes the varying strength of semantic prosody and its probabilistic nature: that a unit with a particular semantic prosody could also be used in a different evaluative environment. As Hunston demonstrates, this is true of the lemma *cause*. Partington also does in fact discuss variation in semantic prosody according to genre, which is identified by Hunston as a point against his view. Given all these arguments advanced by Partington, Hunston’s criticism seems overall to be somewhat excessive. However, it does seem to be the case that as a matter of practice in his published analyses, Partington tends to prioritise considerations of semantic prosody as a feature of an individual item.

Morley and Partington (2009: 144-150) directly address Hunston’s criticism of Partington’s account of semantic prosody. They refute Hunston’s claim by arguing that Hunston’s dichotomy between Sinclair and Partington’s schools of thought, as previously mentioned, is a false one (Morley and Partington 2009: 144). They reason that these two so-called schools are, in fact, “simply two ways of viewing the same phenomenon from different standpoints,” that is, the lexical-priming and the discursal perspectives, and that both Sinclair and Partington accept both perspectives (Morley and Partington 2009: 144). That is, Morley and Partington (2009: 144-148) argue that both Sinclair and Partington accept that semantic prosody is both (a) a psychological phenomenon

regarding the mental links that exist between words, described in terms of lexical priming; and also (b) a discourse phenomenon that can be statistically identified in texts.

It is very clear that Partington does view semantic prosody from both these perspectives. Partington, as we have seen, explicitly bases his discussion of semantic prosody on lexical priming theory, and at the same time, he identifies semantic prosody in texts. But the argument that Sinclair also embraces lexical priming theory is more dubious. Lexical priming is the explicitly psychological concept that every word is primed in our mind to co-occur with other words (McEnery and Hardie 2012: 145; see also section 2.4). But Sinclair does not seem to wish to admit any psychological aspect to his work, as Stubbs argues:

Sinclair remained notoriously unwilling to draw out the social-psychological implications of his work on phraseology. It is only corpora which can provide data for studying prosodies from the bottom up, and therefore show how we could do real ‘ordinary language philosophy’.

(Stubbs 2009: 133)

The above argument thus weakens Morley and Partington’s claim that Sinclair’s conceptualisation of semantic prosody incorporates the psychological aspect that they characterise using Hoey’s lexical priming terminology. In fact, even if we accept that Sinclair adopts the lexical-priming perspective, his discussion of semantic prosody still differs from Partington’s in a number of ways (see section 2.5.2). Therefore, I would agree with Hunston that Sinclair’s approach is in fact different from Partington’s, although Hunston seems clearly to go too far in her implication that Partington has a simplistic view of semantic prosody.

2.3.4.3 Partington's later study of semantic prosody

Recently, it appears that Partington's notion of semantic prosody has undergone some change. His latest concept of semantic prosody emphasises the idea of evaluative meaning, rather than the concept of connotation, as it did before. His focus on evaluation is evident from his suggestion to change the term used for the concept to *evaluative prosody* (Partington 2014: 279), which obviously highlights evaluative meaning.

In his discussion of evaluative prosody, Partington adopts the notion of evaluation proposed by Hunston (2004: 157). Therefore, by evaluation, Partington (2014: 279) means "the indication of whether the speaker thinks that something (a person, thing, action, situation, idea, etc.) is good or bad." He further argues that "goodness and badness can, of course, come in many forms" (Partington 2014: 280). For example, a situation may be positively evaluated because it is pleasurable, or because it is profitable. It may be negatively evaluated because it is dangerous, or because it is difficult.

Proposing this new terminology for semantic prosody, Partington (2014: 283) argues that evaluative prosody can be seen as a lexical item's "inherent potential to participate in evaluative interaction with other items of similar polarity." For example, as Partington argues, due to its positive evaluative prosody (which results from its habitual co-occurrence with positive items), *brimming with* tends to be selected by a speaker who wishes to express that an entity is full of something positive, say *confidence* or *hope*, because the positive evaluative prosody of this phrasal verb has the same evaluative polarity as *confidence* and *hope*. Their combination, say *brimming with confidence*, thus forms consistent positive evaluation in the discourse. On the other hand, a speaker who wishes to express that an entity is full of something bad tends to select *fraught with*,

given its negative evaluative prosody and therefore its potential to combine with other negative words to form consistent negative evaluation. Nevertheless, as Morley and Partington (2009: 146) also note, an item's evaluative prosody may sometimes be switched off or even exploited for stylistic effect (Partington 2014: 283).

Three more points are important in Partington's concept of evaluative prosody. First, evaluative prosody plays a major role in establishing consistent evaluation over longer stretches of discourse. Partington explicitly argues as follows:

Evaluative prosody is perhaps best seen as part of the process or habit of competent speakers/writers of choosing to use together – of co-selecting within a stretch of discourse – items with the same evaluative polarity in order to maintain *evaluative consistency*, that is, what we might call *evaluative harmony*.

(Partington 2014: 284)

Partington elsewhere argues that maintaining evaluative harmony is “very important to avoid sending ‘mixed messages’” (Morley and Partington 2009: 144). Saying *I had to sit through a really exciting concert*, for example, might be confusing for the listener; he or she might wonder if the speaker is being ironic, given fact that *sit through* has a negative prosody, whereas *really exciting* expresses positive evaluation.

Second, studies of evaluative prosody mostly involve items whose evaluative meaning is not immediately apparent from their semantics (Partington 2014: 283). For example, the evaluative potential of *brimming with* and *fraught with* is not evident in the items' basic semantics (both mean “full of, containing many of”). Rather, it is discerned only when we look at their concordances. On the other hand, for items such as *wonderful* and *ridiculous*, the evaluative meaning is already obvious from their semantics, even without the assistance of the concordance.

Third, Partington's evaluative prosody is closely associated with collocation, in that a lexical item can acquire its evaluative prosody from the items it habitually co-occurs with². In turn, an item established to possess a particular evaluative prosody can participate, often through collocation, in establishing evaluation. Partington puts it as follows:

The simplest form of the process we are calling evaluative prosody is seen in collocation, when we see the sharing of evaluative polarity between a node and its collocate.

(Partington 2014: 281)

Despite the change in terminology, the underlying idea of evaluative prosody does not appear greatly different from semantic prosody as it is dealt with in Partington's early work. Just like Partington's version of semantic prosody, evaluative prosody is contingent upon its co-occurring items, and it is still restricted to the positive vs. negative opposition. Evaluative prosody also mostly involves items that are neutral in their basic semantics. This corresponds to the distinction between semantic prosody and expressive connotation that Partington (2004: 131-132) makes in his earlier discussion of semantic prosody. Some of the properties of evaluative prosody noted by Partington are likewise similar to those of semantic prosody, such as its variation in accordance with genre and the possibility for it to be switched off or exploited for stylistic effect.

Partington's discussion of evaluative prosody is therefore arguably a continuation of his version of the notion of semantic prosody. That said, there has been a shift in his argument regarding the location of semantic prosody. Partington earlier consistently argued that semantic prosody spreads across a unit of meaning greater than a word. But

² This notion can arguably be described in Whitsitt's terms as 'meaning transfer', an argument Whitsitt makes to criticise Louw's concept of semantic prosody (see section 2.3.1.3).

this argument does not seem to apply any more, in that he asserts that evaluative prosody is the item's *inherent* evaluative potential to co-occur with other items of the same evaluative polarity. This, one might argue, could lend support to Hunston's earlier argument that Partington primarily treats semantic prosody as a property of an individual item. Nevertheless, even though Partington now explicitly says that evaluative prosody is an inherent feature of a lexical item, he still maintains that Hunston's dichotomy is a false one, supporting his argument with the same rationale presented in his earlier response to Hunston's criticism (see section 2.3.4.2; Morley and Partington 2009: 144-148).

In summary, Partington's study of semantic prosody can arguably be divided into two stages, as evident from his recent suggestion of a change in the terminology for the concept from semantic prosody to evaluative prosody. However, the underlying idea of evaluative prosody is still similar to that of semantic prosody, although the change in the terminology, from my perspective, indicates that Partington has become clear in his view of semantic prosody as an aspect of evaluation. One point that seems to represent a shift in his thinking is his now explicit argument that evaluative prosody is inherent in a lexical item, which as Hunston points out, distinguishes Partington's approach – in both earlier and later work – from Sinclair's. (Partington primarily bases his argument of semantic prosody being inherent in a lexical item on Hoey's lexical priming theory, and also invokes this theory as a rationale to argue against Hunston, as will be discussed in detail in the next section.) Regarding this point, I would agree with Hunston that Partington's approach is indeed different from Sinclair's in a number of respects. This topic will be dealt with in detail in section 2.5.2.

2.4 Lexical priming and semantic prosody

2.4.1 Introduction to lexical priming

Hoey's lexical priming theory primarily concerns the psychological aspect of collocation. Hoey (2005: 7) argues that the pervasiveness of collocation in language can only be explained from a psychological perspective, as "collocation is fundamentally a psychological concept." By this, he means that each word is primed in our minds to co-occur with other words (its collocates) (McEnery and Hardie 2012: 145). Therefore, when we produce a word, we are psychologically primed to produce one of its collocates; likewise, when we perceive a word, we are psychologically primed to expect one of its collocates to follow. Hoey further notes that this knowledge of collocational priming is acquired through the language user's repeated encounters with a word. Specifically, Hoey says:

We can only account for collocation if we assume that every word is mentally **primed** for collocational use. As a word is acquired through encounters with it in speech and writing, it becomes cumulatively loaded with the contexts and co-texts in which it is encountered, and our knowledge of it includes the fact that it co-occurs with certain other words in certain kinds of context.

(Hoey 2005: 8)

Hoey (2004: 386-388) further argues that collocational priming is just one type of priming; there also exist other types, such as grammatical priming, semantic association priming, and colligational priming. Similarly, the knowledge of these types of priming is obtained through the language user's repeated encounters with a word. The grammatical priming of a word underlines the grammatical class that a word belongs to. From the lexical priming perspective, Hoey (2004: 386) argues, instead of saying, for example, that

breakfast is a noun, we should say that it is “primed for use as a noun”. Semantic association occurs when a word or word sequence is “associated in the mind of a language user with a semantic set or class” (Hoey 2005: 24)³. *Hour*, for example, is primed in our mind to co-occur with words belonging to the semantic sets of number and journey, such as *half-hour-drive* and *four-hour-flight* (Hoey 2005: 16). Finally, colligational priming exists when a word or word sequence is mentally “primed to occur in, or avoid, certain grammatical functions or structures” (Hoey 2004: 388). For example, analysing the colligational priming of *consequence*, Hoey (2005: 46-47) discovers that *consequence* tends to occur in the grammatical functions of Adjunct, Complement, and Subject, but to avoid the grammatical function of Object.

Hoey (2005: 9) notes some properties of priming, arguing that these properties apply to all types of priming. First, primings *nest* (Hoey 2004: 389; 2005: 17). The phenomenon of nesting occurs when a word sequence comes to have its own primings (collocational, colligational, or semantic association primings), separately from the primings of the elements that build up the sequence (Hoey 2005: 58). The sequence *young men and women*, for example, is primed to co-occur with words from a semantic set of compliments such as *bright*, *dedicated*, and *finest*, as in *bright young men and women*, and so on. However, neither *men* nor *women* is individually primed for the same semantic association (Bastow 2003 cited in Hoey 2005: 19).

Second, primings are domain and/or genre specific (Hoey 2003: 13). A word or word sequence is primed for use in our minds differently in different types of context. For example, *research* is mentally primed to co-occur with *recent* in academic contexts but

³ Hoey (2005: 24) explicitly argues that his concept of semantic association is the same as Sinclair’s notion of semantic preference (see section 2.3.2.1).

not in other types of context (Hoey 2005: 10). Also, whereas *in winter* is primed to be used in travel writing, *during the winter months* is primed to be used in gardening writing (Hoey 2004: 386).

Third, primings can be neglected or exploited for creativity (Hoey 2004: 387). Even though a word in our minds carries with it a set of primings that, as earlier argued, are acquired through our repeated encounters with it, we are in control of our language and are not obliged to use the word according to how we have been primed (Hoey 2004: 394). Rather, we can ignore the normal primings or override them for creativity. Hoey (2004: 387) refers to the use of *Gift Aid* in the sentence “If your supporter number ends in “D”, you already Gift Aid your donations” that appears in a charity appeal letter as an example to illustrate this point. He argues that here the writer of the letter simply ignores the grammatical priming of *Gift Aid* as a nominal group, but rather uses it as if it was primed for use as a verb group.

Forth, primings may *drift* (Hoey 2005: 9, 11). Particularly, Hoey argues that every time we encounter a word or use it, its priming in our minds is strengthened or weakened. It will be reinforced if what we encounter or choose to produce is familiar to our present priming. By contrast, if what we encounter or choose to produce is unfamiliar to us, our current priming will be loosened, and this will in turn result in a drift or shift in priming, which can occur over the course of our lifetime.

Fifth, primings may *crack* (Hoey 2005: 11, 179). Cracks in priming occur when we receive data about the use of a word or word sequence that are contradictory to our current primings, and we cannot find the way of solving the conflict. The conflicting data may come from self-reflexivity or education. For example, children at school may be told

by their teacher that their use of a particular word is incorrect. Cracks can result in uncertainty about priming, which can in turn cause long-term linguistic insecurity. That said, they can also be mended. For example, we can either reject our current primings or reject the new data we receive. Another better way to fix the crack is that we can assign the current primings to one context, and reserve the new primings for another context (because, as observed earlier, primings are context-bound).

Finally, primings are harmonised and shared (Hoey 2005: 11). It has been earlier argued that the knowledge of priming is obtained through the language user's repeated encounters with a word. Since each language user tends to encounter a word in different various contexts, it then follows that their knowledge of priming should be unique. That said, Hoey (2005: 11; 2004: 392-393) argues that our knowledge of priming is not too vastly different from each other, as there are in fact harmonising principles that "bring our primings into line with each other". These principles are our own self-reflexivity; education, which is the most important; shared literary and religious tradition; the mass media; and dictionaries and grammars. Due to such harmonising principles, we come to share a number of primings and are able to communicate with each other, despite having different experiences of language.

2.4.2 What lexical priming adds to semantic prosody

We have seen that lexical priming theory primarily concerns the psychological aspect of language use. Particularly, a word is primed in our minds to be used in specific contexts. Morley and Partington express this in terms of a set of instructions of use:

Thus, although we might hesitate to talk of lexical items having intrinsic context-free meaning, they do carry with them a set of instructions on how to use them, on how they normally interact with other items, that is to say, they have, in Hoey's terminology, *primings*. Among these is the semantic prosody of an item, which is information or a suggestion, which tells the reader "when you find me instantiated in a text, you are likely to find me in combination with items of a like, consistent evaluative force..."

(Morley and Partington 2009: 145)

So we see that, as Morley and Partington argue, a word's primings also include information about its semantic prosody, that is, which evaluative patterns, positive or negative, a word is primed in our minds to occur within. This information, Partington (2014: 288) argues, is not obvious to introspection, even though we are intuitively aware of how to make use of it.

I observed in section 2.3.4.1 that Partington bases his discussion of semantic prosody on lexical priming theory⁴. Taking this point further, we can see that some notions from lexical priming theory could serve as an explanation for some properties of semantic prosody noted by Partington as follows.

Hoey argues that primings can be switched off (see section 2.4.1). This property of primings could arguably account for the probabilistic nature of semantic prosody noted by Partington (see section 2.3.4.1). To be more specific, this concept could explain why a lemma such as *bent on*, as Partington observes, can occur in a positive environment without indicating irony (as Louw suggests it must). Re-expressed in terms of lexical priming, we can say that even though *bent on* is primed in our minds to co-occur with negative words, we do have the ability to choose to ignore that priming and instead use it in positive environments.

⁴ Partington has confirmed in personal communication that he adopts lexical priming theory in his discussion of semantic prosody/evaluative prosody.

Hoey contends that primings are genre specific (see section 2.4.1). This notion is arguably in line with Partington's notion that semantic prosody varies across genres (see section 2.3.4.1). Partington, for example, argues that *lavish* displays a negative semantic prosody in newspaper prose, but tends to display a positive semantic prosody in the entertainment field (see section 2.3.4.1). Re-expressed in Hoey's terms, we can say that *lavish* is mentally primed to occur in negative contexts in newspaper prose, but primed to occur in positive environments in the entertainment field. This property of priming could in fact help solve the problem Hunston argues for in Partington's discussion of semantic prosody. Particularly, it could account for the counter-examples Hunston finds regarding the lemma *cause* (see section 2.3.4.2). In fact, in her discussion of the counter-examples, Hunston (2007: 263) states that "it would be possible to suggest that this verb [the lemma *cause*] loses its association with negative evaluation when it occurs in 'scientific' registers." Hence, in Hoey's terms, we can say that the lemma *cause* does not display its typical negative prosody because it has different primings in the scientific genre as opposed to other genres.

Partington not only employs the theory of lexical priming in his discussion of semantic prosody, but also, as we have seen, invokes it as part of his rationale when arguing against Hunston that his approach to semantic prosody is in fact not different from Sinclair's (see section 2.3.4.2). Particularly, as we have seen, Partington argues following Hoey (2005: 13) that a word's priming also includes, alongside information about its collocational and grammatical use, information about its semantic prosody (Morley and Partington 2009: 146). Therefore, he contends that for him to talk of a word as "having" a semantic prosody, a point Hunston identifies as separating him from

Sinclair, is actually “useful shorthand” for saying that a word carries with it a particular set of instructions, or primings in Hoey’s terminology (Morley and Partington 2009: 145). He goes on to add that in fact several scholars – including Sinclair himself and Hunston – also talk of semantic prosody in this way.

Thus, we see that lexical priming theory plays an important part in Partington’s conception of semantic prosody. Other scholars, such as Louw, Sinclair and Stubbs do not, however, adopt Hoey’s lexical priming theory in their discussion of semantic prosody, although many similar ideas are clearly present in their accounts.

2.5 Discussion of similarities and differences among the approaches to semantic prosody

2.5.1 Similarities among the approaches to semantic prosody

One feature that seems common to almost all approaches to semantic prosody proposed in the literature is its evaluative or attitudinal quality (Stewart 2010: 21). The attitudinal function of semantic prosody is prioritised by Sinclair (2004: 34) when he defines semantic prosody as being “attitudinal, and on the pragmatic side of the semantics/pragmatics continuum” (see section 2.3.2.1). This quality is also emphasised by Stubbs, not unexpectedly considering the fact that Stubbs’ idea of semantic prosody, especially in his later work, operates within Sinclair’s framework (see section 2.3.3.2).

Partington (1998: 66; 2004: 131) also stresses the evaluative function of semantic prosody. He defines the evaluative function of semantic prosody in connotative terms, arguing that it is an aspect of expressive connotation (see section 2.3.4.1). This function is also argued to be primary – it is what semantic prosody is *for*. As Morley and

Partington (2009: 143) put it, semantic prosody “is the way in which speakers establish and maintain *connotational* or *evaluative harmony* within a stretch of discourse by co-selecting items of a consistent evaluative/attitudinal force.” For instance, *setting in* tends to get selected over, say, *beginning* or *commencing* in *rotten weather seems to be setting in*, because, unlike the neutral to good prosodies of the other two verbs, the negative prosody of *set in* is consistent with *rotten*, which is evidently negative.

It could be argued that Sinclair and Stubbs also associate semantic prosody with the idea of connotation as well. Sinclair (2003: 117) argues that a word often carries with it “extra meaning of an emotive or attitudinal nature” in a particular context. He goes on to argue that this kind of meaning is sometimes called “connotation” or “pragmatic” meaning. As he puts it:

...this kind of meaning is structurally important, and essential for the understanding of language text. We will call it SEMANTIC PROSODY – semantic because it deals with meaning, and prosody because it typically ranges over combination of words in an utterance rather than being attached to just to one.

(Sinclair 2003: 117)

Stubbs also seems to consider semantic prosody or discourse prosody (in his terminology) analogous or even equivalent to connotation (Stewart 2010: 27). This can be inferred from Stubbs’ (2001: 96) discussion of the elements of extended lexical units, which he also refers to as semantic schemas, when he writes:

These semantic schemas can be modelled as clusters of lexis (node and collocates), grammar (colligation), semantics (preferences for words from particular lexical fields) and pragmatics (connotations or discourse prosodies).

(Stubbs 2001: 96)

In fact, the status of semantic prosody as a synonym of connotation is apparent in Stubbs' discussion of the negative semantic prosody of the lemmas *cause* and *affect*, as we have seen in section 2.3.3.1.

Louw also makes reference to the evaluative function of semantic prosody. As with the other scholars mentioned, Louw's (2000: 50) discussion of semantic prosody touches upon the idea of connotation, but in contrast to them, he argues that "semantic prosodies are not merely connotational." Specifically, Louw distinguishes semantic prosody from connotation by arguing that the former is "more strictly functional and attitudinal" and "relate[s] more directly to what literary critics call authorial *tone*", while "knowledge of connotations is often a form of schematic knowledge of repeatable events". Whitsitt (2005: 286), however, finds Louw's argument unconvincing, reasoning that authorial tone is "precisely how connotation has often been described."

Despite the fact that almost all approaches to semantic prosody mention evaluative or attitudinal function, evaluation may in fact not *always* be present (Stewart 2010: 22). For instance, Stewart (2010: 23) considers the use of *set in* in the idiom "rigor mortis set in", rather than, say, *started* or *began*. Since *set in* is almost always used here, the use of *set in* seems to be practically automatic. If this is the case, the item's ability to "bring to bear any extra pragmatic nuance is severely restricted" (Stewart 2010: 23). Thus, it follows that the assumption that a negative attitude is expressed in *rigor mortis set in* is questionable. In cases of highly fixed usages, then, semantic prosody may in fact not serve an evaluative function.

2.5.2 Differences among the approaches to semantic prosody

We have seen that semantic prosody has been variously approached in the literature. Interestingly, most scholars assert that they adopt Sinclair's approach to semantic prosody. Louw (1993: 158) credits Sinclair for having provided him with both the term and the concept. Stubbs (2001: 87) clearly states in his discussion of a model of extended lexical units, of which semantic prosody is one of its element, that he develops "proposals in Sinclair 1996, 1998." Partington (2004: 132) implies that he adopts Sinclair's approach when he writes that semantic prosody "spread[s] over a unit of language which potentially goes well beyond the single orthographic word and is much less evident to the naked eye." This is an endorsement of an idea that originates with Sinclair (2004: 33), when he comments, regarding the semantic prosody of *naked eye*, that the prosody is shown by a combination of a word such as *small* or *faint* with *see* (see section 2.3.2.1), and notes elsewhere that semantic prosody is "subject to enormous variation, making it difficult for a human or a computer to find it reliably" (Sinclair 1998: 20).

It can, nevertheless, be argued that in reality these scholars do not completely adhere to Sinclair's approach. Of the three scholars, I would argue that Stubbs' understanding of the concept of semantic prosody seems to be the closest to Sinclair's, in that he argues in favour of Sinclair's proposal that semantic prosody is a compulsory element of an extended unit of meaning. There appear, however, to be some differences in terms of their practice. Louw and Partington's approaches, by contrast, are different from Sinclair's in terms of both the concept and the practice. These latter two approaches are, nonetheless, quite similar to each other, although some points of difference can also

be discerned. In this section, I will primarily consider comparisons and contrasts across four areas: how semantic prosody is identified; whether semantic prosody is only positive or negative; where semantic prosody resides; and whether diachronic or synchronic considerations are involved in the concept of semantic prosody.

2.5.2.1 How semantic prosody is identified

There are differences in the four scholars' practice in terms of identifying semantic prosody. Louw and Stubbs seem to adopt a similar approach, in that they identify a lexical item's semantic prosody by focusing on individual words found in the concordance lines around that lexical item. Louw identifies semantic prosody from an item's collocates, as explicitly stated in his definition of semantic prosody (see section 2.3.1.1). Particularly, Louw identifies collocates through concordance reading. His dependence upon collocates in identifying semantic prosody is apparent in his identification of the negative prosody of *utterly* when he writes:

If we study the collocates to the right of *utterly*, we find a phenomenon similar to that identified for *set in*. The concordance shows that *utterly* has an overwhelmingly 'bad' prosody: there are few 'good' right-collocates.

(Louw 1993: 160)

Stubbs also primarily identifies semantic prosody from individual co-occurring words. In his early study, he identifies semantic prosody based on statistical collocates (as I outlined in see section 2.3.3.1). Stubbs continues to identify semantic prosody based on collocates in his later study, even after he adopts Sinclair's idea of the model of extended unit of meaning. However, this time, Stubbs identifies collocates through concordance reading (see section 2.3.3.2). For example, in his investigation of the lemma

undergo, Stubbs (2001: 89) examines the lemma's concordance lines in order to identify the first 20 most frequent collocates. Having identified these collocates, Stubbs (2001: 90) goes on to argue that the "involuntary" prosody is usually realised to the left of *undergo* by the collocates *forced* and *required*, which appear in the top 20 collocates list. The prosody is also realised by the collocate *must*, which is "one of the most frequent collocates amongst the stop-words" (Stubbs 2001: 90).

It is worth noting in this context that an issue has been raised regarding the application of the term *collocate* in the identification of an item's semantic prosody. Stewart (2010: 86) argues that it is generally accepted that the term *collocation* refers to *habitual* co-occurrence, but the use of the term *collocate* in the analysis of semantic prosody does not seem to be consistent with this concept. For example, as Stewart points out, Louw does not give any suggestion as to whether the co-occurrence between the node and a particular lexical item (which Louw refers to as *collocate*) is habitual, frequent, and/or statistically significant. Therefore, Louw's use of the term *collocate* seems to refer to simple co-occurrence, not necessarily habitual co-occurrence (Stewart 2010: 86).

Stubbs' application of the term *collocate*, by contrast, appears to be different from Louw's. By *collocates*, Stubbs is referring to habitual co-occurrence. This is clear from the fact that he suggests employing statistic methods, and in fact himself employs statistic methods to identify collocates in his early study of semantic prosody. In his later study, even though he changes to identify collocates through concordance reading rather than relying on statistic methods, he does mention the frequency of co-occurrence. In this thesis, like Stubbs, I will reserve the term *collocate* for habitual co-occurrence.

Partington also seems to identify semantic prosody from individual co-occurring words, which he calls “collocates”, in his early study of semantic prosody. For instance, using *commit* as an example of an item displaying an unfavourable connotation (or semantic prosody), Partington (1998: 66-67) argues that the unfavourable prosody resides in the word *commit* as well as its unpleasant collocates such as *offences*, *deception*, and *suicide*. It seems likely that here Partington’s (1998) use of the term *collocate* also refers to simple co-occurrence, not habitual co-occurrence, because as with Louw, he does not mention frequency or statistical significance.

Elsewhere, Partington (2004: 133-144) identifies the semantic prosody of a word using what he calls the word’s “environment”, particularly in his analyses of the *happen* word group (see section 2.3.4.1). For example, discussing the prosody of *occur*, Partington (2004: 141) writes: “These results appear largely analogous to those for HAPPEN, with over twice as many bad environments as good ones.” While Partington does not give a completely precise account of what he means by *environment*, it can be deduced from his discussion of his analyses that *environment* refers to two things. First, it can refer to an individual co-occurring word or a co-occurring phrase, identified through concordance reading. For example, Partington (2004: 142) states that the lemma *come about* “co-occurs regularly with the following items denoting *processes*: *change(s)* (3 times), *assimilation*, *adaptation* and *process* itself (twice), as well as with whole phrases which outline some process or other”. One thing that should be noted is a change in his terminology. In his early studies, Partington uses the term *collocate* to refer to individual co-occurring words. However, in this later work, it appears that he changes to use the term *item* instead. Second, the term *environment* can mean stretches of text. For instance,

in his analysis of *happen*, Partington gives one example of the bad events that contribute to the negative semantic prosody of the lemma as follows:

She then went to bed feeling more relaxed, but after a while felt unwell and was soon sick. This happened several times during the night.

(Partington 2004: 136)

It can be seen that *this*, the subject of *happened*, refers to the whole preceding stretch of text, rather than to any individual co-occurring word or phrase.

In Partington's later work on semantic prosody, after his move to using the term *evaluative prosody*, he appears to continue to identify the evaluative prosody of a word using both individual co-occurring words and longer stretches of text. In his discussion of *fraught with*, Partington identifies the negative evaluative prosody of the phrase based on individual co-occurring negative items such as *danger/s*, *risk/s* and *peril*. Again, in this study, Partington uses the term *item* rather than *collocate* to refer to such individual co-occurring words. In his analysis of *par for the course*, by contrast, Partington relies on stretches of text. He argues that the negative items that contribute to the negative evaluative prosody of the phrase, as identified by Channell (2000 cited in Partington 2014: 282), can sometimes be "entire stretches of text, which can even appear across clause or sentence boundaries" and gives an example as follows:

George Bush is talking again and I don't have a clue what he's saying. It's not that he is *mangling his syntax*. That's **par for the course**.

(Partington 2014: 282)

Likewise, Partington (forthcoming) identifies the positive evaluative prosody of *make a difference* from stretches of text, arguing that the phrase "is very generally found in positive contexts, even though there appears to be nothing explicitly positive or

negative in the semantics of the expression itself.” One of his examples of the use of *make a difference* in positive contexts is as follows:

A number of vacancies are on offer that could provide a real chance to **make a difference** to youngsters’ lives and further your career.

(Partington forthcoming)

It can be concluded at this point that Louw and Stubbs adopt quite similar methods for identifying semantic prosody. They tend to identify semantic prosody from individual co-occurring words, which they call “collocates”, although Louw means by this anything that co-occurs, while Stubbs means only habitually co-occurring words. Whereas Louw identifies what he calls “collocates” only through concordance reading, Stubbs sometimes uses statistical measures and sometimes uses concordance reading. Like Louw and Stubbs, Partington on some occasions identifies semantic prosody based on individual co-occurring words, which he refers to as “collocates” in his early work, but more generally as “items” in his later studies. That said, Partington does not restrict himself to simply looking at individual co-occurring words identified through concordance reading. He also identifies semantic prosody based on stretches of text or wider contexts. We have seen from his analyses of *par for the course* and *make a difference* that the phrases’ negative and positive evaluative prosodies respectively arise from a contextual expression of evaluation across more than one word. Stewart (2010: 87), commenting on the inconsistent use of the term *collocate* for simple versus habitual co-occurrence, seems to argue in favour of Louw’s usage when he posits that semantic prosody is the phenomenon of co-occurrence regardless of whether the relationship between the items that co-occur is habitual or significant or not.

By contrast, Sinclair identifies semantic prosody by observing concordance lines, and looking for pragmatic functions expressed over an extended co-text, rather than by looking at particular individual words (frequent or otherwise) (McEnery and Hardie 2012: 138). For example, analysing *naked eye*, Sinclair (2004: 33) argues that the prosody of “difficulty” of *naked eye* “may be shown by a word such as *small, faint, weak, difficult* with *see*” – that is by the combination of both words. To put it another way, the semantic prosody of “difficulty” need not be evident in any individual word, but rather can be expressed pragmatically over a string of words (see section 2.3.2.1).

Therefore, we see that Louw, Stubbs and Partington are broadly similar in their approach to identifying semantic prosody. Sinclair, by contrast, adopts a different method. Interestingly, Louw, Stubbs and Partington assert that their approach is similar to Sinclair’s or follows Sinclair’s. In reality, what they do is different from what Sinclair did, as has been illustrated.

2.5.2.2 Whether semantic prosody is only positive or negative

Louw and Partington generally restrict semantic prosody to being either positive or negative, although Partington also uses the terms *favourable* and *unfavourable*. However, even though both scholars describe semantic prosody similarly, there are other differences in their understanding of the phenomenon. Under Louw’s view, if a lexical item has been assigned a particular prosody, whether positive or negative, its use in a context that does not fit its prosody is indicative of either irony or insincerity (see section 2.3.1.2). By contrast, Partington’s view is that the prosody of an item is gradable and that

an item being used in atypical environment does not necessarily *always* indicate irony or insincerity (see section 2.3.4.1).

Sinclair, on the other hand, does not confine himself to the positive vs. negative opposition (Stewart 2010: 11). Rather, as has been seen, he explores a broader co-text to identify the pragmatic function of an extended unit of meaning, which can be any function/meaning, not just positive/negative. For example, he identifies “difficulty with visibility” to be the prosody of the unit whose core is *naked eye*, and assigns the prosody of “frustration in the face of difficulty” to the extended unit around *budge*, to name but two.

Stubbs appears inconsistent regarding this point. In his early study, he restricts himself to the positive vs. negative opposition, as in his analysis of the lemma *cause* (see section 2.3.3.1). But later, he does not confine semantic prosody to being only positive or negative. Rather, semantic prosody can express any pragmatic functions of an extended unit of meaning, as can be seen in Stubbs’ discussion of the “involuntary” prosody of the unit whose core is the lemma *undergo* (see section 2.3.3.2). This is probably because in this later work he is operating more clearly within Sinclair’s framework of semantic prosody.

So, we see that while Louw and Partington limit semantic prosody to the positive vs. negative opposition, Sinclair and, in his later studies, Stubbs do not. However, Partington argues that in fact his restriction to the positive vs. negative opposition is just “a two-term Linnaean-style binomial notation” (Morley and Partington 2009: 141). This good-bad parameter can actually represent a number of forms of goodness and badness. Given this, Morley and Partington (2009: 141) assert that their use of the positive vs.

negative opposition is not dissimilar to Sinclair's practice, as positive can be read as "pleasurable", or "profitable", and negative as "difficult" or "dangerous", and many more. This argument does not come as unexpected, considering that Partington has consistently argued that his approach to semantic prosody is very alike to Sinclair's. However, I would argue that Morley and Partington's argument is not strong enough to support their claim. Despite their explanation of "positive" and "negative" as representing many other more specific concepts, I would argue that their practice and idea of semantic prosody are still different from Sinclair's for two reasons. First, the more specific concepts that positive or negative evaluative prosodies may represent are still limited to expressions of good or bad evaluation, even if the topic of evaluation can be varied. Sinclair, by contrast, does not restrict semantic prosody to varieties of the expression of evaluation. Rather, semantic prosody can be any pragmatic function or meaning, although it is obvious from his analyses that the semantic prosodies he identifies for *naked eye*, *true feelings*, and *budge* are in fact associated with evaluation (see sections 2.3.2.1, 2.5.2.3 and 2.3.2.2). Thus, I do not think that an approach that is substantively oriented around the good-bad parameter can avoid being classified as substantially different from Sinclair's approach in at least that respect. A second contrast that persists between Partington's style of semantic prosody analysis and Sinclair's is that the specific evaluation that the good-bad parameter may represent is, in Partington's analyses, derived from the analysis of semantic preference. Partington (forthcoming), for example, argues that *fraught with* habitually co-occurs with three types of negativity: "danger/risk, negative emotions, difficulty/complications and negative emotions". Based on these *semantic preferences* (see section 2.3.3.2), Partington goes on to argue that *fraught with*

has a negative evaluative prosody, further explaining that “negative” here can be read as “dangerous”, “difficult, over-complex”, and “negative feelings”. It is clear from this analysis then that the negative evaluative prosody that can be read as “dangerous, etc.” is derived from the phrase’s *semantic preferences*. Sinclair, by contrast, clearly distinguishes methods for identifying semantic prosody from methods for identifying semantic preference, as will be discussed later.

The method of looking at individual co-occurring words is used, by all researchers, to identify semantic preference (see section 2.3.3.2 for Stubbs’ definition and an example of semantic preference). In all cases, to identify semantic preference, a more detailed analysis of semantic category, not just positive or negative, is required. Given that the method of looking at individual co-occurring words is adopted by some scholars, in particular Stubbs and Partington, to identify both semantic prosody and semantic preference, it is understandable these scholars find it hard to clearly distinguish between the two concepts. Stubbs notes that:

The distinction between semantic preference and discourse prosody is not entirely clear-cut. It is partly a question of how open-ended the list of collocates is: it might be possible to list all the words in English for quantities and sizes, but not for “unpleasant things”.

(Stubbs 2001: 66)

Partington asserts that:

Semantic prosody is a sub-category, or a special case, of semantic preference, to be reserved for instances where an item shows a preference to co-occur with items that can be described as bad, unfavorable or unpleasant, or as good, favourable or pleasant.

(Partington 2004: 149)

We thus see that if we identify semantic prosody from individual co-occurring words, the distinction between semantic prosody and semantic preference is likely to be

fuzzy. However, this fuzziness tends to disappear if we adhere to Sinclair's method for identifying semantic prosody. Sinclair draws a sharp distinction regarding the methods for identifying these two different phenomena. That is, whereas he focuses on (categories of) individual co-occurring words to identify semantic preference, he looks for functions expressed pragmatically in the extended unit around the core to identify semantic prosody (see section 2.3.2.1). Hence, following Sinclair's method, the overlap between semantic prosody and semantic preference is much reduced, thus removing the fuzziness of the distinction between the two phenomena. (Bednarek 2008: 129-133 attempts to distinguish semantic preference from semantic prosody on other grounds, but not, in my view, successfully⁵.)

There thus seems to be a connection between the method used to identify semantic prosody and whether the prosody is restricted to the positive vs. negative opposition. That is, if semantic prosody is mainly identified from individual co-occurring

⁵ Bednarek (2008: 121) argues that semantic preference is a collocational phenomenon. She further argues that there are two types of collocational phenomenon: the POS/NEG collocation and the SEM collocation. Whereas the POS/NEG collocation refers to cases where a word shows a tendency to co-occur with positive or negative words, the SEM collocation refers to cases where a word tends to co-occur with words belonging to a particular semantic group. So we can see that the phenomenon that Bednarek labels as POS/NEG collocation and argues to be one of the two types of semantic preference is generally regarded as *semantic prosody* by most other scholars. Her concept of semantic preference is, then, arguably confusing from the start. With regard to semantic prosody, Bednarek (2008: 132) argues that, following Sinclair, she reserves the term semantic prosody for "connotations of all kinds" including the POS/NEG connotation of a single word. But though Sinclair (2003: 117) also associates semantic prosody with the idea of connotation (see section 2.5.1), the central feature of semantic prosody for Sinclair is its pragmatic function that is spread across an extended unit of meaning rather than being evident from a single particular word of the unit (see section 2.3.2.1). Moreover, although other scholars such as Louw, Stubbs and Partington also associate semantic prosody with connotation, Bednarek's equation of semantic prosody and connotation is rather more strongly asserted than are those of these scholars (see section 2.5.1). Therefore, I would argue that Bednarek's redefinitions confuse rather than clarify the different ideas she discusses.

words, it is typically restricted to the positive vs. negative opposition – because if the classification of the individual co-occurring words is not confined to this opposition, what will be obtained from the analysis will just be semantic preference. But if semantic prosody is derived from an extended co-text, it can be expressive of any pragmatic function or meaning, broad or narrow, and still be a phenomenon clearly distinct from semantic preference.

2.5.2.3 Where semantic prosody resides

The issue of where semantic prosody resides has recently been extensively debated in the study of semantic prosody. According to the literature, there are two main views: one is that it resides in an individual item alone, and the other is that it belongs to an item and its co-text (or an extended unit of meaning, in Sinclair’s terminology) (Stewart 2010: 57).

Louw does not explicitly address this point. Nevertheless, his definition of semantic prosody as an aura of meaning of a form seems to suggest that semantic prosody belongs to a lexical item alone. This notion can also be implied from the phrasing he uses to describe semantic prosody. By way of illustration, when he writes: “The concordance shows that *utterly* has an overwhelmingly ‘bad’ prosody,” he appears to suggest that he views the negative semantic prosody as a feature or property of *utterly*. Likewise, if we consider the binary claim Louw makes about prosodic clashes, we could argue that if he did not view semantic prosody as a property of the item, he would not be able to make this claim. It would not be possible to claim that prosodic clashes are *always*

indicative of either irony or insincerity unless semantic prosody is considered to be a feature that an item carries with it all the time.

Partington (2004: 132) makes explicit his view on semantic prosody as a feature of an item and its co-text when he writes that semantic prosody “spread[s] over a unit of language which potentially goes well beyond the single orthographic word.” That said, it could be argued that the phrasing he uses to describe semantic prosody in practice may contradict this conceptualisation to some extent. As Stewart (2010: 58) points out, when Partington comments on the connotation of the word *commit*, he argues that the unfavourable connotation of *commit* is expressed not only by the word *commit* alone, but also by its collocates. This, Stewart argues, gives the impression that he considers semantic prosody a property of an extended unit. However, immediately after this remark, Partington goes on to comment on *rife*, arguing that it has an unfavourable semantic prosody as an item. For Stewart, this latter comment clearly contradicts Partington’s former remark on *commit*. In fact, as has been mentioned, there are numerous occasions when Partington’s phrasing gives the impression that semantic prosody is a property of a lexical item alone, for example when he writes that the lemma *set in* “has an extremely unfavourable prosody” (Partington 2004: 135). Even if we set aside Partington’s phrasing, it can still be argued that he prioritises the notion of semantic prosody as a feature of the lexical item alone, as argued by Hunston (2007: 250; see also Stewart 2010: 58). (See section 2.3.4.2 for Hunston’s argument and Partington’s defense against Hunston’s critiques.) In fact, recently, suggesting the change in the terminology for the concept from semantic prosody to evaluative prosody, Partington (2014: 283) explicitly argues that evaluative prosody is an inherent feature of a lexical item (see

section 2.3.4.3). As I have argued above, this shift can be seen as clearly distinguishing his approach from Sinclair's.

Sinclair, on the contrary, explicitly argues that semantic prosody is an obligatory component of an extended unit of meaning; it does not reside in any particular item in the unit but rather is pragmatically spread across the whole unit. This is shown, for example, in his commenting on the collocation *true feelings* that it is “the core of a compound lexical item” that has “a semantic prosody of reluctance/inability” as its inherent component (Sinclair 2004: 35).

Stubbs (1995: 1), in his early study of semantic prosody, views semantic prosody as a property of a lexical item, as apparent in his remark at the very beginning in one of his papers that “words have distinctive semantic profiles or ‘prosodies’”. He later operates under Sinclair's framework of the model of an extended lexical unit and thus embraces the idea of semantic prosody as belonging to that unit, as has been seen in his discussion of the lemma *undergo* (see section 2.3.3.2).

Thus, it could be argued that Louw and Partington's concepts of semantic prosody are similar, in that they primarily view semantic prosody as a feature of an individual item alone. This view contrasts with that of Sinclair and Stubbs, in his later study, who argue that semantic prosody is spread across an extended unit of meaning.

2.5.2.4 Whether diachronic or synchronic considerations are involved in the concept of semantic prosody

This issue has been extensively discussed by Stewart (2010: 41-48). His major argument is that any definition of semantic prosody that involves the notion that semantic prosody is a meaning that an item *acquires* is diachronic in nature. That is, to define

semantic prosody as the result of meaning transfer is to suggest that it is the result of a process that takes place over time.

Louw's description of semantic prosody, as we have seen in section 2.3.1.1, clearly involves the concept of meaning transfer: the meaning of the "collocates" is transferred to the node. Recall that by the term *collocate*, Louw means simple, not habitual or frequent, co-occurrence, as discussed in section 2.5.2.1. A similar idea is embraced by Partington. He explicitly invokes the concept of meaning transfer in his argument that a (node) item does not *always* acquire the evaluative sense of its "collocates", as discussed in section 2.3.4.1. Therefore, we might argue following Stewart that Louw and Partington conceptualise semantic prosody within a diachronic framework.

This conceptualisation of semantic prosody is found to be problematic by Stewart (2010: 47). He argues that a description of the phenomenon that suggests diachronic transfer of meaning is not consistent with "the way the *analyses* are carried out", which tends to be synchronic in nature. Particularly, he points out that analyses of semantic prosody are mostly restricted to synchronic analyses, as reflected in the use of synchronic corpora, whereas "diachronic *analyses* are entirely absent in the literature" (Stewart 2010: 47). This point is also raised by Whitsitt, who emphasises the problems in using synchronic corpora for the analysis of semantic prosody that is diachronically framed, as we have seen from his attacks against Louw (see section 2.3.1.3).

Sinclair, by contrast, primarily identifies semantic prosody in synchronic terms (Stewart 2010: 11). This is evident from his major argument that semantic prosody is an obligatory element of an extended unit of meaning that expresses the pragmatic function

of the unit (see section 2.3.2.1). This suggests that semantic prosody or pragmatic meaning can be considered an inherent part of the unit of meaning, rather than something that has necessarily arrived from somewhere else.

Following Sinclair's approach, Stubbs, in his later studies, also identifies semantic prosody within a synchronic framework. In his earlier studies, nevertheless, we have seen that he closely links semantic prosody to collocation and suggests that an item's prosody is determined by its collocates (see section 2.3.3.1). Thus, as with Louw and Partington, his early concept of semantic prosody implicitly involves the idea of meaning transfer and is thus implicitly diachronically framed, even though the corpora he uses in his analyses consist only of synchronic data.

2.5.3 Summary of similarities and differences among the approaches to semantic prosody

We have seen that the approaches to semantic prosody proposed in the literature share one key similarity, but otherwise differ in a number of ways. In terms of similarity, we have seen that the attitudinal or evaluative quality of semantic prosody is a feature embraced by all four scholars. Sinclair puts an emphasis on the attitudinal or pragmatic function of semantic prosody as a compulsory element of an extended unit of meaning. This notion is also embraced by Stubbs in his later discussion of the phenomenon. Partington stresses that semantic prosody is an aspect of expressive connotation that conveys evaluative meaning. Louw, like Partington, refers to the concept of connotation and in particular emphasises the role of semantic prosody in establishing authorial tone.

There are, however, notable differences among these scholars' approaches to semantic prosody. We have seen that even though Louw, Stubbs and Partington argue

that they follow Sinclair's approach, there are in fact points where their approaches can be seen to differ from Sinclair's. The differences among the four scholars' approaches to semantic prosody are summarised in Table 2.2.

Scholars	Louw	Sinclair	Stubbs	Partington
Points of difference				
How semantic prosody is identified	Individual co-occurring items	Pragmatic meanings expressed across extended units of meaning	Individual co-occurring items	Individual co-occurring items/ contextual expressions of evaluation
Whether semantic prosody is only positive or negative	Yes	No	Yes (early work)/ no (later work)	Yes
Where semantic prosody resides	Individual lexical items	Extended units of meaning	Individual lexical items (early work)/ extended units of meaning (later work)	Individual lexical items
Whether semantic prosody is diachronic or synchronic in nature	Implicitly diachronic	Synchronic	Diachronic (early work)/ synchronic (later work)	Diachronic

Table 2.2 Differences among the four scholars' approaches to semantic prosody

We can see from Table 2.2 that Louw and Partington's approaches are very similar. Their approaches are, however, considerably different from Sinclair's. Stubbs is somewhat inconsistent in his approach; although he argues that he follows Sinclair's approach in his later work, and his definition of the phenomenon clearly has moved closer to Sinclair's, I would argue that his later approach still inclines towards Louw and Partington's, rather than towards Sinclair's, in terms of his actual technique for

identifying semantic prosody. In fact, the practice of looking at individual co-occurring items adopted by Louw, Stubbs and on some occasions Partington is, as Hoey (2005: 24) contends, used by Sinclair to identify semantic *preference*, not semantic *prosody* (see section 2.5.2.2). This argument of Hoey's helps reinforce the point that Sinclair's approach to semantic prosody is indeed different from the other three scholars'.

From all the above discussions, I would thus argue that in general there are two prevailing approaches to semantic prosody in the literature. The first approach is represented by the studies of Louw, Stubbs, and Partington. Within this approach, semantic prosody is mainly identified from individual co-occurring words ("collocates"). There are, however, certain differences between each scholar's practice. For example, whereas by the term *collocate* Louw (and, generally, Partington) means simple co-occurrence, Stubbs means habitual co-occurrence. Moreover, while Stubbs sometimes uses statistical measures, Louw and Partington identify collocates through concordance reading only. Basing his analysis on concordance reading, Partington also on some occasions identifies contextual expressions of evaluation that extend over longer stretches of text than a single word. Within this approach, semantic prosody tends to be restricted to the positive vs. negative opposition. Even though Partington argues that goodness and badness can come in many forms, I would argue that his practice and notion of semantic prosody is still dissimilar to Sinclair's (see section 2.5.2.2). Finally, within this approach, semantic prosody is viewed as residing in an individual item (often a single word), and is conceptualised within a (implicitly or explicitly) diachronic framework.

The second approach is represented by the work of Sinclair. Within this approach, semantic prosody is identified from pragmatic meanings that are expressed over an

extended co-text. It is not confined to the positive vs. negative opposition, but can be any pragmatic function or meaning. It is an inherent part of unit of meaning, and is framed synchronically.

These are the two approaches that I will employ in my analysis. In this thesis, I will label the approach represented by the studies of Louw, Stubbs, and Partington the *polarity-oriented approach*, as its central feature is that semantic prosody is restricted to positive vs. negative evaluative polarity (sometimes interpreting this, as Morley and Partington put it, as “binomial notation” for various good/bad oppositions). I will label Sinclair’s approach the *EUM-oriented approach*, where *EUM* stands for “extended unit of meaning”. Within this approach, semantic prosody can be any pragmatic function or meaning, and is not confined to the positive vs. negative opposition. It must be noted that within the polarity-oriented approach, although I will in general use the three scholars’ method for identifying semantic prosody, restrict semantic prosody to the positive vs. negative opposition, and view semantic prosody as a property of a word, I do not adopt the notion that semantic prosody is a result of meaning transfer over time. I agree on this point with Whitsitt and McEnery and Hardie that we cannot make claims about a diachronic transfer of meaning based on observations derived from synchronic corpora (see section 2.3.1.3). Therefore, within this approach, I view semantic prosody instead as a synchronic phenomenon, that is, as a tendency of a word to occur in positive or negative contexts.

Stewart (2010: 161) builds a similar argument to this in his discussion of Sinclair and Louw’s approaches. However, he does not consider Stubbs and Partington’s approaches in his discussion. Stewart ultimately argues that there are two primary

approaches to semantic prosody in the literature: Louw's approach and Sinclair's approach. Whereas Louw views semantic prosody as a feature of the word and considers it within the diachronic framework, Sinclair views it as a compulsory element of an extended unit of meaning within the synchronic framework. Stewart's conclusion is thus compatible with, though less comprehensive than, the characterisation of the field at which my review of the literature has arrived.

2.6 Studies of semantic prosody in English

Semantic prosody has been extensively explored in English. However, since my thesis focuses on the study of semantic prosody in another language, I will give only a brief overview of literature on semantic prosody in English. Table 2.3 gives some examples of the English expressions whose semantic prosody has been studied in the literature.

Scholars	Positive semantic prosody	Negative semantic prosody
Partington (1998: 66-74)	PROVIDE	COMMIT
Stubbs (1996: 174)		Rife
Hunston (2002: 60-62)		PEDDLE
Schmitt and Carter (2004: 8)		Pedlar/peddler
Stewart (2010: 3)		Dealings
		SIT THROUGH
		Bordering on
		BREAK OUT

Table 2.3 Examples of semantic prosodies in English

The table appears to confirm Louw's (1993: 171) argument that there seem to be more negative semantic prosodies than positive ones, although it could also be that people are more interested in studying negative semantic prosodies. Of all the examples,

only the lemma *provide* has a positive semantic prosody. These scholars appear to adopt the same approach in their study. They identify semantic prosody from individual co-occurring words and restrict semantic prosody to the positive vs. negative opposition (see section 2.5.2.1 and section 2.5.2.2). The majority of studies of semantic prosody in English adhere to this approach. However, some studies instead operate within Sinclair's framework. One is the study of Tognini-Bonelli.

Tognini-Bonelli (2001: 106-110) analyses the extended unit of meaning whose core is the adjective *proper* in attributive position. Examining the right co-text of *proper*, Tognini-Bonelli identifies collocates such as *consideration*, *consultation*, *medicine*, and *peace*. These collocates, she argues, share a semantic feature: they are all approved things that normal citizens should have. Therefore, the semantic preference for approvables is established. In the left co-text, Tognini-Bonelli notices the idea of absence, realised by words such as *no*, *not*, and *never* at the grammatical level and by expressions such as *prevent or delay*, *failing to provide*, and *without giving* at the lexical level. This idea of absence, coupled with the semantic preference of approvables, points to the semantic prosody of “a ‘complaint for the absence of something that we all think should be present or available’”, which is the function of the unit of meaning which has *proper* as its core (Tognini-Bonelli 2001: 110).

The studies mentioned so far analyse semantic prosody as a phenomenon in its own terms. However, the notion of semantic prosody is also exploited in a variety of areas such as (critical) discourse analysis, stylistics, and sociolinguistics. Mautner's (2007) research on the representation of “the elderly” is an example of such a study in the area of discourse analysis. Mautner (2007: 51-58) argues that, due to the negative

semantic prosody of *elderly*, which results from its habitual co-occurrence with negative items such as *infirm*, *disabled*, *frail*, *victims*, etc., “*elderly* is primarily associated with discourses of care, disability, and vulnerability.” Another example in the area of critical discourse analysis is Baker’s (2006: 73-92) investigation of discourses of refugees. Examining the concordance of *refugee/refugees*, Baker (2006: 84-87) finds that these terms co-occur frequently with words suggesting tragedy, such as *plight* and *despair*, and in patterns characterising refugees as criminal or a nuisance. On the basis of the prosodies, Baker therefore concludes that the terms “refugee” and “refugees” are associated with the discourses of victims and criminals/nuisances.

We have seen so far that all the studies discussed in this section except Tognini-Bonelli’s adopt Louw, Stubbs, and Partington’s approach. One reason for the predominance of this approach could be that it is easy and efficient, especially when the analysis of semantic prosody is just part of a larger study, as in Mautner or Baker’s studies. Studies of semantic prosody in other languages also seem to be dominated by this approach, as will be discussed in detail in the next section.

2.7 Studies of semantic prosody in other languages

Semantic prosody has also been explored in languages other than English, although fairly few. In fact, apart from Chinese, only European languages such as Portuguese, Spanish, Danish, Norwegian and Italian have been studied in this way. Studies of semantic prosody in these languages often feature a comparison of semantic prosodies in English and the language in question, although the methodology adopted may be different. In this section, I will discuss the studies of Berber Sardinha (2000),

Munday (2013), Dam-Jensen and Zethsen (2008), Xiao and McEnery (2006), Wei and Li (2013), Ebeling (2013; 2014), and Tognini-Bonelli (2002).

2.7.1 Berber Sardinha

Berber Sardinha (2000: 93) analyses the semantic prosodies of the Portuguese translations of the English (phrasal) verbs *commit* and *set in* in a Portuguese corpus. The analysis demonstrates that the pair *commit/cometer* (the only translation suggested by the *Dicionário Inglês-Português*, on which the identification of translations is based) shares a negative semantic prosody. The negative prosody of *cometer* is evident in the verb's negative collocates such as *crime/s*, *erro/s* 'errors', and *faltas* 'fouls/faults'. The picture is more complicated regarding *set in*, where four Portuguese translations are identified: *manifestar-se*, *estabelecer-se*, *entrar*, and *cair*. The verb *manifestar-se* has a negative semantic prosody only when it means 'affect' and collocates with words such as *doença* 'disease/sickness', whereas the other three translations do not at all display a negative semantic prosody. Berber Sardinha therefore concludes that *set in* has no direct translation-equivalents in Portuguese.

Generally, Berber Sardinha (2000: 106) concludes that "semantic prosodies may vary across Portuguese and English." He also points out the insufficiency of dictionaries, arguing for the inclusion of semantic prosody information in dictionaries, which are regular resources for translators.

2.7.2 Munday

Munday (2013) investigates the semantic prosodies of the lemma *loom large* and one of its Spanish dictionary correspondents *cernerse*. His analysis of *loom large* in the BNC shows that it is overwhelmingly used in negative environments, in particular with the subjects referring to “ominous ‘landscape’” such as *dark water* and to “metaphorical negative of failure, upheaval and prospect” like *chaos* and *trade war*. Like *loom large*, *cernerse* is found in the CREA Spanish Real Academic corpus to be typically used in negative contexts, though not always. The subjects of *cernerse* can be categorised into nature subjects, meteorological phenomena, metaphorical ominous meteorology, and subjects expressing dangers, threats, and uncertainty, with *amenaza(s)* ‘threat(s)’ being the most common collocate.

Munday (2013: 179) concludes that *loom large* and *cernerse* “have generally negative semantic prosodies.” But he points out that the verb phrases differ in terms of their most common collocates and the syntactic patterns where they occur. For example, while *loom large* is usually followed by the prepositional phrase *in ... mind* such as *in her mind*, this pattern is absent from *cernerse*, which regularly precedes the pattern *sobre* ‘over’ + place/people/group of people. In addition, whereas *amenaza(s)* is the most common collocate of *cernerse*, typically realised in the pattern *amenazas* + relative pronoun + present tense of *cernerse* + *sobre* + noun, its English translation, *threat(s)*, does not co-occur at all with *loom large*.

Munday (2013: 181) suggests complementing this kind of contrastive study, which is primarily based on dictionaries and monolingual corpora, with a study of actual translations in translated texts. He further contends that cooperation between translation

studies theorists, corpus linguists, and software developers could considerably enhance research on semantic prosody.

2.7.3 Dam-Jensen and Zethsen

Dam-Jensen and Zethsen (2008) explore the semantic prosodies of two Danish verbs *forårsage* ‘cause’ and *medføre* ‘lead to’ in a Danish corpus, Korpus 2000. The analysis demonstrates that *forårsage* has a strong negative semantic prosody, with all of the observed instances in the concordance occurring in negative environments. This finding confirms their hypothesis that *forårsage*, being a translation-equivalent of *cause*, would likewise display an extremely negative semantic prosody. *Medføre* also displays a negative semantic prosody, but its negative prosody is weaker than that of *forårsage*, as it also occurs in neutral or even positive context, and the effects implied by *medføre* are less serious and less likely to happen (as evident from the use of hedging) than those implied by *forårsage*.

Dam-Jensen and Zethsen (2008: 1620) conclude that *forårsage* exemplifies what they call *inherent evaluative meaning*, where its prosody is so strong that the (negative) evaluation should be considered part of its semantic meaning. On the other hand, *medføre* exemplifies what they call *pragmatic evaluative meaning*, where the expression has a clear (negative) evaluative tendency, but this tendency is not strong enough to form part of its meaning and thus prevent it from occurring in neutral or positive context.

2.7.4 Xiao and McEnery

Xiao and McEnery (2006) investigate the collocational behaviour and semantic prosodies of three groups of near synonyms in English and Chinese: the *consequence* group, the *cause* group, and the *price/cost* group, mainly basing their analysis on FLOB, Frown and the Lancaster Corpus of Mandarin Chinese.

Four near synonyms are identified for the *consequence* group: *result*, *outcome*, *consequence*, and *aftermath*. The analysis demonstrates that these synonyms “can be arranged, from positive to negative, on a semantic continuum as follows: *outcome/result*, *consequence*, and *aftermath*” (Xiao and McEnery 2006: 111). Turning to the Chinese equivalents (identified from a bilingual dictionary), *jie2guo3*, like its equivalents *outcome/result*, “typically does not express a negative affective meaning” (Xiao and McEnery 2006: 112). *Hou4guo3*, the equivalent of *consequence* and *aftermath*, is habitually negative, although, like *consequences*, it expresses a neutral affective meaning when used in academic prose.

In the *cause* group, it is found that *cause* is typically negative. Its near synonyms, e.g. *bring about*, *result in/from*, *lead to*, *arouse* and *give rise to*, display a different semantic prosody. While *bring about* is predominantly positive, the rest can be positive or negative. The Chinese equivalents vary in terms of semantic prosody. Some are overwhelmingly negative, e.g. *zhi4shi3* and *yin3fa1*, while others are positive, e.g. *cu4shi3*.

Finally, for the *price/cost* group, Xiao and McEnery find that the semantic prosody is contingent upon the patterns the item occurs in and whether it is used literally or metaphorically. For instance, when *price* and *cost* are used metaphorically in the

patterns *at a price* and *at a cost*, they convey a negative affective meaning. But when they are used in a literal sense, they typically denote a neutral meaning. Their Chinese equivalent, *dai4jai4*, is found to denote a neutral meaning when used literally, but a negative meaning when used metaphorically, which is its typical use. Its near synonyms, e.g. *jia4ge2* and *jia4qian2*, by contrast, are all used in a literal sense, conveying a neutral meaning.

Xiao and McEnery (2006: 124-125) conclude that the collocational behaviour and semantic prosody of near synonyms in both English and Chinese are different, so near synonyms cannot normally be used interchangeably. The results also show that semantic prosody may vary according to domains, and that the collocational behaviour and semantic prosody of near synonyms and their close translation-equivalents can be similar as well as different. Finally, Xiao and McEnery (2006: 125) point out the pedagogic implications of the study. For example, due to possible differences in semantic prosody, teachers should be careful when explaining vocabulary using near synonyms. Rather, they are encouraged to present vocabulary in context so that learners can observe both its meaning and usage.

2.7.5 Wei and Li

Wei and Li (2013) explore the semantic preferences and semantic prosodies of four translation-equivalent pairs across Chinese and English. Their method for identifying the equivalent pairs differs from that of the previously discussed scholars, though, in that they identify the equivalent pairs on the basis of a bi-directional parallel corpus (the Shanghai Jiao Tong University Parallel Corpus), rather on a bilingual

dictionary. However, the concordance analyses of the specific terms are still based on comparable monolingual corpora (the Chinese National Corpus and the British National Corpus).

Wei and Li (2013: 105) argue that semantic prosody does not reside in a single word, but rather in “a particular phraseological patterning of a word”. Therefore, a word, they further argue, “may be associated with more than one prosody⁶, realized in more than one pattern of co-selection, and that cross-linguistic equivalence only resides in corresponding patterns of co-selection of words under study” (Wei and Li 2013: 105). A term, *Major Patterning*, is coined to refer to the most frequent pattern of co-occurrence, and *Minor Patterning* to refer to a less frequent one.

For reasons of space, I will discuss in detail only one translation-equivalent pair studied by Wei and Li: *yongxian/spring up*. Their analysis of *yongxian* shows that the noun subjects of *yongxian* have a semantic preference for newness/excellence. Of 100 occurrences of *yongxian*, 98 indicate a positive attitudinal meaning, which means, Wei and Li argue, that the verb has a strong positive semantic prosody. Turning to the English equivalent, *spring up*, the analysis shows that its noun subjects have a semantic preference for new institutions/entities and abstractions/mental processes. These semantic preferences in turn contribute to two characteristic types of patterning. One is the Major Patterning, comprising “a positive attitudinal meaning, a semantic preference of newness, and the node item”, and the other is the Minor Patterning, containing “a negative attitudinal meaning, a semantic preference of abstractions/mental processes, and the node item” (Wei and Li 2013: 116). The overall analyses thus show that there are both similarities and differences between *yongxian* and *spring up*. Both display a semantic

⁶ This contrasts with Louw’s binary claim (see section 2.3.1.2).

preference for newness and express a positive semantic prosody; yet *yongxian* has a stronger positive prosody. Two other pairs, *yongxian/crop up* and *ganran/infect* demonstrate divergent semantic preferences and prosodies. By contrast, the pair *cehua/orchestrate* displays a convergent semantic preference and prosody.

Wei and Li (2013: 134) conclude that the nature and strength of semantic prosody of a translation pair are “important indicators of degree of equivalence or non-equivalence.” If the items of a pair exhibit a convergent strong prosody, this generally indicates a higher degree of equivalence.

2.7.6 Ebeling

Ebeling (2013) studies the semantic prosodies of *cause* as a noun and a verb and its Norwegian translation-equivalents. Like Wei and Li, Ebeling identifies the translation-equivalents from a bidirectional translation corpus (the English-Norwegian Parallel Corpus), but unlike Wei and Li, Ebeling also uses this parallel corpus for her main analysis.

Ebeling first identifies the semantic prosody of *cause* (both as a verb and a noun) by scrutinising concordance lines with *cause* in both the English originals and English translations. The analysis shows that *cause* displays a negative semantic prosody, even though “the negative prosody is more clearly represented when *cause* is used as a verb” (Ebeling 2013: 7). She then identifies the translations of *cause* in the Norwegian translated texts and examines their semantic prosodies in the Norwegian originals. There emerge three main verb correspondents. The results demonstrate that *få (x til å)* ‘get (x to)’, the most frequent equivalent, “does not have the same degree of negative semantic

prosody as *cause*” (Ebeling 2013: 9), while *føre til* ‘lead to’, the third most common equivalent, is closest to *cause* in terms of semantic prosody. The second most frequent correspondent, *forårsake* ‘cause/bring about’, appears only three times in the Norwegian texts, so Ebeling is unable to comment on it. Three main noun equivalents are identified. *Årsak* ‘cause’, which is most commonly used, is closest in terms of semantic prosody to *cause*. *Grunn* ‘reason’, the second most frequent correspondent, is largely found in neutral contexts, although it also occurs in negative environments.

Ebeling concludes that there is a gap between English and Norwegian. Even though there exist clear counterparts of *cause* in Norwegian, they are not real equivalents, because they “do not have the same degree of negative semantic prosody as *cause*” (Ebeling 2013: 11). This gap is reaffirmed in Ebeling’s (2014) study, where she adopts the same technique to examine the semantic prosodies of the Norwegian translation-equivalents of *commit*, *sign of*, and *utterly*. The analysis shows that *utterly* resembles *cause*, in that it does not seem to have a good Norwegian match in terms of negative semantic prosody, although good matches emerge for *commit* and *sign of*. One thing worth noting is Ebeling points out that when she refers to the semantic prosodies of *cause*, *commit*, *sign of*, and *utterly*, she means the semantic prosodies of the units whose core is *cause*, *commit*, *sign of*, and *utterly*, respectively. This thus means that Ebeling is directly referencing Sinclair’s conceptualisation, even though it appears, in particular in her 2014 study, that she primarily looks at individual co-occurring words when identifying semantic prosody.

We have seen that in all the studies discussed so far semantic prosody is largely identified from individual co-occurring words or collocates and is restricted to the

positive vs. negative opposition. Hence, these contrastive studies are consistent with the approach adopted by Louw, Stubbs, and Partington, as discussed in section 2.5.3. We saw earlier that Tognini-Bonelli's (2001) English study operates within Sinclair's framework. Her subsequent contrastive study (Tognini-Bonelli 2002) works within this framework as well, as will be discussed in the next section.

2.7.7 Tognini-Bonelli

Tognini-Bonelli (2002) analyses the formal patterning of three pairs of translation-equivalent expressions across English and Italian: *in the case of/nel caso di*, *in case of/in caso di*, and *in case/se per caso* in several monolingual corpora (the Economist corpus, the Wall Street Journal corpus, the Birmingham corpus, and an unnamed Italian corpus).

Tognini-Bonelli (2002: 91) here proposes the notion of “functionally complete units of meaning”. She argues following Sinclair that units of meaning are larger than a single word, because “words do not live in isolation, but in strict semantic and functional relationship” (Tognini-Bonelli 2002: 91). To put it another way, words often occur in patterns, and when these patterns of co-selection are obvious and strong, they form “functionally complete units of meaning” (Tognini-Bonelli 2001: 106). She goes on to argue that units of meaning will be functionally complete when their collocation, colligation, semantic preference, and semantic prosody are identified, and only when they are functionally complete “do they become available as translation equivalents” (Tognini-Bonelli 2002: 91-92). Tognini-Bonelli's concept of a “functionally complete unit of meaning” is therefore arguably equivalent to Sinclair's extended units of meaning.

I will discuss only one equivalent pair studied by Tognini-Bonelli: *in the case of/nel caso di*. The expression *in the case of* displays colligation with the definite article *the* to the right. No strong collocational patterns can be identified, however. Its semantic preference varies: “people alternate with countries, tangible objects with less tangible ones” (Tognini-Bonelli 2002: 84). This semantic preference is expressed by proper names such as *Australia, the Liberals, and London’s motorways*. The prevalence of the article *the*, coupled with a proper name, points to the semantic prosody of “the introduction of specificity”. The translation-equivalent *nel caso di* also colligates with the definite article. Like *in the case of*, no strong collocational patterns can be identified for *nel caso di*. The expression has a semantic preference for two areas: “the area of technical and scientific terminology”, expressed by nominals such as *acciai superferritici* and *algoritmicarcinoma midollare*⁷, and “the area of literary analysis”, expressed by nominals such as “*Lord Spleen*” and *l’Orlando Furioso*. Like its English equivalent, *nel caso di* has a semantic prosody of specificity. Tognini-Bonelli concludes that equivalence is therefore established between *in the case of* and *nel caso di* at the level of functionally complete units of meaning, despite the differences in semantic preference.

Finally, Tognini-Bonelli (2002: 91) argues in favour of the approach adopted in her study as “a way of establishing and evaluating the comparability of units of meaning across languages”, which could considerably benefit contrastive and translation studies. She also emphasises the importance of comparable corpora as an indispensable tool in this kind of analysis.

In addition to the contrastive study discussed above, Tognini-Bonelli (2001: 113) studies the Italian verb *andare incontro*, which means “going toward someone to meet

⁷ Tognini-Bonelli does not provide the English translation of these nominals.

them”, in its own terms. She finds that this verb, which is often used metaphorically, displays a negative semantic prosody, with an idea of unpleasantness frequently indicated by the noun phrases that precede or follow the verb. This investigation, though interesting, is comparatively brief.

Many scholars who work with semantic prosody might find Tognini-Bonelli’s analysis of *in the case of/nel caso di* unusual, or perhaps not even a semantic prosody at all. One reason would be that the semantic prosody of introduction of specificity that Tognini-Bonelli identifies does not express any attitudinal or evaluative meaning, which is the central feature of semantic prosody that all the four main scholars discussing the phenomenon embrace (see section 2.5.1). It should be recalled, in this context, that Tognini-Bonelli’s idea of a functionally complete unit of meaning is equivalent to Sinclair’s idea of an extended unit of meaning, and that they adopt the same approach to semantic prosody. Thus, her identification of the introduction of specificity as the semantic prosody of *in the case of/nel caso di* could arguably be seen as straying from Sinclair’s indicated approach, as it is not associated with attitude or evaluation. However, I would argue the contrary: that Tognini-Bonelli *is* in fact in line with Sinclair’s approach. Even though Sinclair identifies the attitudinal function as the central feature of semantic prosody (see section 2.3.2.1), and all the semantic prosodies that he identifies appear to incorporate attitude or evaluation, as we have seen with *naked eye*, *true feelings*, and *budge* (see sections 2.3.2.1, 2.5.2.3 and 2.3.2.2), the *pragmatic function* is also key to Sinclair’s characterisation of semantic prosody, as evident from the following statement defining semantic prosody, which has already been introduced in section 2.3.2.1:

It [semantic prosody] expresses something close to the ‘function’ of the item – it shows how the rest of the item is to be interpreted functionally. Without it, the string of words just ‘means’ – it is not put to use in a viable communication.

(Sinclair 2004: 34)

Thus, I would argue that introducing a semantic prosody of the introduction of specificity is not unusual within, or a departure from, Sinclair’s notion of semantic prosody, even if it is not associated with attitude or evaluation. Tognini-Bonelli (2002: 84) in fact explicitly argues that she is “not associating a particular evaluation” with *in the case of*. Rather, the introduction of specificity can be seen as “the ultimate function of this complex preposition”. That is, as Tognini-Bonelli (2002: 83) argues, its function is to “present individual examples, considered for their particular characteristics.” So, here the introduction of specificity is the pragmatic function of the extended unit of meaning whose core is *in the case of*; the speaker’s desire to express that function is, to use another Sinclair’s formulations, the motivation for the use of the unit. Moreover, this unit colligates with *the*, and has a semantic preference for people, countries, tangible and less tangible objects. Overall, the semantic prosody is spread across the whole sequence of “in the case of + the + proper name” rather than evident from any individual item. Scholars who adopt the polarity-oriented approach may well view such a claimed semantic prosody as unusual or even wrong. However, as I have now argued, this is an analysis entirely in line with Sinclair’s stated approach to semantic prosody oriented around the concept of the extended unit of meaning.

A semantic prosody analysis of this kind may potentially invite one further kind of criticism. It might be objected that the semantic prosody identified through this approach can end up being simply a paraphrase of the semantics of the core of the

extended unit of meaning. For example, it could be argued that the notion of “introduction of specificity” is in fact a paraphrase of one of the meanings of *case*, which can mean ‘a specific example’, and so this kind of analysis is of limited use – or, to put it more strongly, perhaps even pointless and not worth doing at all. But from within the perspective that motivates the analysis to begin with, that is the EUM-oriented approach, even semantic prosodies which are high-paraphrases of the core do additionally illustrate a central tenet of the Sinclairian theory of language: words do not appear randomly, but rather, particular (meanings of) words tend to appear in fairly narrowly-defined lexical, grammatical and semantic contexts (see sections 2.2.2 and 2.3.2.1) – and we have seen exactly this with Tognini-Bonelli’s analysis of *in the case of*. That is to say, even though the semantic prosody that she identifies could be seen as a paraphrase of the core of the unit, her analysis additionally discovers that *in the case of* does not occur randomly. Rather, it tends to co-occur with the article *the*, not *a* or *an*, and with proper names. Thus, I would argue that the Sinclairian approach to semantic prosody still merits investigation, as it forms a coherent approach given the basic ideas of the theory of language it is anchored within, despite the criticisms that might be made from other perspectives.

2.7.8 Summary of studies of semantic prosody in other languages

We have seen that all of the studies discussed, with an exception of Tognini-Bonelli’s small study of *andare incontro*, are contrastive studies between English and a given language. Three important points of difference can be noted. First, while some studies concentrate on English items previously established to have a particular prosody and explore their translation-equivalents in another language, others select items that

have never been studied before in any language. Second, the methodology adopted varies across the studies, particularly in terms of the type of corpora on which the analyses are based. Third, all the studies, except that of Tognini-Bonelli, adopt Louw, Stubbs, and Partington's approach. Tognini-Bonelli proposes the idea of functionally complete units of meaning, which is equivalent to Sinclair's model of an extended unit of meaning. She often identifies semantic prosody from pragmatic meanings expressed over an extended co-text, rather than from individual co-occurring words, and does not restrict her analyses to the positive vs. negative opposition, as other scholars do. In spite of these differences, generally the results of the studies show both similarities and differences in terms of semantic prosody between English and the language under study. These similarities and differences may in turn have implications for the degree to which the expressions being compared are truly translation-equivalents.

2.8 Concluding remarks

This chapter has discussed the literature relevant to the research questions of the thesis. I first provided some brief background on neo-Firthian corpus linguistics. Then I presented a detailed discussion of the approaches to semantic prosody existing in the literature, as well as some criticisms of these approaches. This included a review of Louw, Sinclair, Stubbs, and Partington's approaches. We have seen within this review some proposals for modified terminology to refer to the concept of semantic prosody. Stubbs suggests the terms *pragmatic prosody* or *discourse prosody* (see section 2.3.3.2); Partington proposes the term *evaluative prosody* (see section 2.3.4.3); and Bublitz, as mentioned earlier, suggests *emotive prosody* (see section 1.1). In this thesis, I will

continue to use the term *semantic prosody*, purely because this is the original terminology that was first used to label the phenomenon. After the review of existing approaches to semantic prosody, I then discussed the similarities and differences among these approaches. This discussion led me to argue that in general there are two primary approaches to semantic prosody. The first approach is represented by the studies of Louw, Stubbs, and Partington, and the second approach is represented by the studies of Sinclair. Whereas the first approach has been extensively used to study semantic prosody in English and many other languages (see sections 2.6 and 2.7), the second approach has not been as extensively used as the first approach. It has been employed only by Sinclair and Tognini-Bonelli to study semantic prosody in English and Italian (see sections 2.3.2 and 2.7.7). Neither approach has been adopted to examine semantic prosody in Thai. Therefore, in Chapter 4, which is the first analysis chapter, both approaches will be applied to the Thai data so as to evaluate their effectiveness and advantages and disadvantages. The results that will be obtained from Chapter 4 will be considered when choosing the optimal approach for the remaining two analyses. Finally in this chapter, I reviewed earlier studies of semantic prosody in English and other languages, including contrastive studies. Having reviewed the relevant literature, in the next chapter, I will present the methodology and the data that will be employed in the thesis.

Chapter 3 – Methodology

3.1 Introduction

In this chapter, I will introduce the corpora to be utilised in the analysis, which are the Thai National Corpus (in section 3.2) and the British National Corpus (in section 3.3). I then move on to discuss the methodology. Here I discuss only aspects of my methods that are general across the whole study. Other aspects that are specific to each analysis will be covered in the relevant chapter. Particularly, the two primary approaches to semantic prosody, which are Louw, Stubbs, and Partington's approach and Sinclair's approach, will be detailed (in section 3.4). These are the two primary approaches which the previous chapter's review of the literature identified. It is necessary to outline how these two primary approaches will be employed in the study, because my first analysis (in Chapter 4) aims to evaluate the effectiveness and the advantages and disadvantages of these two approaches (research question 1). The results obtained from this first analysis will be considered when selecting the most appropriate approach for each of the remaining two analyses, which will address research questions that concern variation in semantic prosodies across genres (research question 2) and similarities and differences in semantic prosodies of Thai-English translation-equivalents (research question 3).

3.2 The Thai National Corpus

The Thai National Corpus (TNC) is a general corpus of present-day standard Thai (Aroonmanakun 2007: 4; see also section 1.3.1). The construction of the corpus was one of the projects to celebrate the anniversary of King Bhumipol's 80th birthday in 2007, under the patronage of H.R.H. Princess Maha Chakri Sirindhorn. It is being developed by the Department of Linguistics, Chulalongkorn University, with collaboration from a number of researchers and publishers and from IBM Thailand. The TNC will become a crucial resource for not only researchers working in the areas of corpus-based lexicography, Thai language teaching, and Thai language processing, but also ordinary people who wish to explore the complexity of the Thai language (Aroonmanakun 2007: 5).

The TNC is not the only corpus in the Thai language that exists. Others reported in the literature are ORCHID, TaLAPi and BEST. ORCHID is the first part-of-speech tagged Thai corpus (Aw *et al.* 2014: 125; Sornlertlumvanich *et al.* 1999: 189), and was constructed by the Communications Research Laboratory of Japan, in collaboration with the National Electronics and Computer Technology Centre of Thailand (Sornlertlumvanich *et al.* 1999: 189). It consists of 400,000 words, all drawn from academic texts (Isahara *et al.* 2000: 466). TaLAPi is another part-of-speech tagged corpus that consists of 4,224,958 words: 1,043,471 from the entertainment and lifestyle domain and 3,181,487 from the news domain (Aw *et al.* 2014: 125). The BEST corpus contains about 5,000,000 words drawn from four different genres: academic articles, encyclopaedia entries, novels and news (Kosawat 2009 cited in Aw *et al.* 2014: 125). We therefore see that, in comparison to the TNC, which at present contains around 32 million

words and covers a wide range of genres (which will be discussed later), these corpora are relatively small and are limited in their genre coverage. They are thus unlikely to provide as much data on as wide a range of semantic prosodies as the TNC, and the TNC is therefore the corpus most likely to generate usable insights relative to my research aims. In what follows, I provide a brief overview of the TNC, which is drawn from Aroonmanakun (2007) and Aroonmanakun *et al.* (2009) unless otherwise indicated.

3.2.1 Structure of the TNC

The TNC is designed to be comparable to the British National Corpus (BNC) (see section 3.3), although at present only written texts are included in the corpus. The criteria for text selection are similar to those of the BNC in terms of domain and medium. In terms of domain, the TNC aims to have 75% informative texts and 25% imaginative texts. The reason for the greater proportion of informative texts is that it is believed by the corpus creators that people in general read or write more informative texts (Aroonmanakun 2007: 6-7). In terms of medium, the TNC plans to have 60% of its texts from books, 20% from journals and newspapers, 5-10% from other published works such as brochures and leaflets, 5-10% from unpublished works (namely letters and notes), and about 5% from texts on the Internet. Since the corpus aims to represent present-day standard Thai, its creators plan to sample 90% of the texts from the period 1998 to 2007 and only 10% from the period prior to 1997. Table 3.1, adapted from Aroonmanakun (2007: 7), summarises the weights of domain, medium, and time in the TNC.

Domain	Weight	Medium	Weight
Imaginative	25%	Book	60%
Informative	75%	Periodical	20%
Applied science		Published miscellanea	5-10%
Arts		Unpublished miscellanea	5-10%
Belief and thought		Internet	5%
Commerce and finance			
Leisure		Time	Weight
Natural and pure science		1998-2007	90-100%
Social science		1988-1997	0-10%
World affairs		Before 1988	0-5%

Table 3.1 Summary of weights of domain, medium, and time in the TNC (after Aroonmanakun 2007: 7).

Texts in the TNC are categorised into genres based on external factors, such as the purpose of communication, the participants, and the setting of communication. This idea is adopted from Lee's proposal for categorising texts in the BNC into genres based on external factors. Lee (2001: 54-55) proposes a genre categorisation scheme for the texts in the BNC, which includes 70 genres: 46 for written texts and 24 for spoken texts. Lee (2001: 55) assigns each BNC text to a genre. Table 3.2 gives an overview of the genre categorisation used in the TNC, which its creators adapted from Lee (2001: 57-58).

Genres	Sub-Genres
<i>Academic</i>	<i>Humanities, e.g. Philosophy, History, Literature, Art, Music</i>
	<i>Medicine</i>
	<i>Natural Sciences, e.g. Physics, Chemistry, Biology</i>
	<i>Political Science – Law – Education</i>
	<i>Social Sciences, e.g. Psychology, Sociology, Linguistics</i>
	<i>Technology & Engineering, e.g. Computing, Engineering</i>
<i>Non-Academic</i>	<i>Humanities</i>
	<i>Medicine</i>
	<i>Natural Sciences</i>
	<i>Political Science – Law – Education</i>
	<i>Social Sciences</i>
	<i>Technology & Engineering</i>
<i>Administration</i>	
<i>Advertisement</i>	
<i>Biography</i>	
<i>Commerce – Finance – Economics</i>	
<i>Religion</i>	<i>(not philosophy)</i>
<i>Institutional Documents</i>	
<i>Instructional – DIY</i>	
<i>Law & Regulation</i>	
<i>Essay</i>	<i>School</i>
	<i>University</i>
<i>Letter</i>	<i>Personal</i>
	<i>Professional</i>
<i>Blog</i>	
<i>Magazine</i>	
<i>Newspaper</i>	<i>Editorial – views</i>
	<i>Agriculture news</i>
	<i>Crime news</i>
	<i>Economic news</i>
	<i>Education news</i>
	<i>Entertainment news</i>
	<i>Foreign news</i>
	<i>Local news</i>
	<i>Politics news</i>
	<i>Sciences & Technology news</i>
	<i>Society news</i>
	<i>Sports news</i>
	<i>Royal family news</i>
	<i>Miscellaneous</i>
<i>Fiction</i>	<i>Drama</i>
	<i>Poetry</i>
	<i>Prose</i>
	<i>Short Stories</i>
<i>Miscellanea</i>	

Table 3.2 An overview of the genre categorisation used in the TNC (from Aroonmanakun 2007:9-10)

3.2.2 Text Sampling

The sample of a text to be included in the TNC is randomly chosen from either the beginning, the middle, the end, or many parts of the whole text. The size of the sample text varies, but does not exceed 40,000 words.

3.2.3 Encoding the TNC

The TEI (Text Encoding Initiative) is used as the markup system of the corpus. Three types of information are marked in the document using XML: documentation on the encoded data (i.e. text metadata); primary data (basic elements in the text, such as paragraphs, and sentences); and linguistic annotation. In terms of linguistic annotation, word boundaries and Latin-alphabet transcription are marked for every word, in the style `<w tran="kot1maaj4">กฎหมาย</w><w tran="thaN3">ทั้ง</w>`. Word segmentation is of use because in Thai texts, word boundaries are not indicated, and this could affect search results. For example, unless words are segmented, a search for รก ‘messy’ might yield not only the word รก, but also undesired results containing that two-character sequence, such as แทรก ‘insert’, เกษตรกร ‘farmer’, or นรก ‘hell’. The TNC Tagger program is used for segmenting words, transcribing pronunciation, and marking texts with `<w>` and `<p>` tags. As I will discuss later (section 6.5.2), my own work has in fact identified certain limitations in the corpus’s automatic word segmentation.

3.2.4 The TNC at present

The TNC is designed to consist of at least 80 million words. However, at present only approximately 32 million words have been added to the TNC. To make the corpus

accessible to the public, a web interface to search the TNC, the TNC web⁸, has been developed. This web interface allows the user to generate concordances, collocations, and distribution data. Searches can be restricted according to domain, genre, year of publication, authors' age, and authors' gender.

3.3 The British National Corpus

The British National Corpus (BNC) is a general corpus of contemporary British English (Aston and Burnard 1998: 28; Hunston 2002: 15). It consists of approximately 100 million words, 90% of which are written texts and 10% of which are spoken texts from a wide range of genres (Aston and Burnard 1998: 28-31; Leech *et al.* 1994: 622). The construction of the BNC, which was sponsored by the Science and Engineering Research Council and the UK Department of Trade Industry, was a collaboration of many institutions, including Oxford University Press and Lancaster University (Leech *et al.* 1994: 622). The following brief overview of the BNC is drawn from Aston and Burnard (1998) unless otherwise indicated.

3.3.1 Structure of the BNC

The written texts in the BNC were chosen according to three criteria: domain, time, and medium. In terms of domain, the BNC consists approximately of 75% informative texts and 25% imaginative texts. With regard to medium, 59% were drawn from books, 31% from periodicals, 4% from miscellaneous published texts (including brochures, leaflets, manuals, and advertisements), 4% from miscellaneous unpublished texts (including letters, memos, reports, minutes, and essays), 1.5% from written-to-be-

⁸ <http://www.arts.chula.ac.th/~ling/TNCII/>

spoken texts (including scripted television material and play scripts), and 0.4% from other unclassified texts. Informative texts were sampled from 1975 onwards, but imaginative texts were sampled from 1960 onwards. Table 3.3 summaries the weights of domain, medium, and time in the BNC.

Domain	Weight	Medium	Weight
Imaginative	21.91%	Book	50.58%
Arts	8.08%	Periodical	31.08%
Belief and thought	3.40%	Miscellaneous published	4.38%
Commerce and finance	7.93%	Miscellaneous unpublished	4%
Leisure	11.13%	To-be-spoken	1.52%
Natural and pure science	4.18%	Unclassified	0.40%
Applied science	8.21%	Time	Weight
Social science	14.80%	1960-1974	2.26%
World affairs	18.39%	1975-1993	89.23%
Unclassified	1.93%	Unclassified	8.49%

Table 3.3 Summary of weights of domain, medium, and time in the BNC (after Aston and Burnard 1998: 29-30)

The spoken texts in the BNC consist of informal encounters and formal encounters. The informal encounters are unscripted conversations recorded by respondents who were recruited based on demographic criteria, such as age, sex, social class, and geographic region. These recordings resulted in 4.2 million words of unscripted conversational English. By contrast, the formal encounters are meetings, debates, lectures, etc., which were selected based on topic and type of interaction. 6.1 million words of the corpus stems from these latter recordings.

3.3.2 Text sampling

Most text samples in the BNC range from 40,000 to 50,000 words, although there exists considerable variation in size. In particular, spoken conversations are quite long, whereas some written unpublished texts, such as school essays, are very short.

3.3.3 Encoding and annotating the BNC

Most recent release of the BNC uses the Text Encoding Initiative's standardised XML markup to encode textual structure and bibliographic information as metadata.

The BNC has been part-of-speech tagged by the CLAWS4 system. CLAWS (Constituent-Likelihood Automatic Word-Tagging System) is "a system for tagging English language texts" by assigning each word in a text to a grammatical class (Garside 1987: 30).

3.4 Methodology

Earlier in this thesis, I argued there are two primary approaches to semantic prosody (see section 2.5.3). Both approaches will be employed in this thesis as follows.

3.4.1 Louw, Stubbs, and Partington's approach (the polarity-oriented approach)

3.4.1.1 General considerations

Louw and Stubbs heavily rely on collocates in their identification of the semantic prosody of a word. Whereas Louw means by *collocates* any co-occurrence, habitual or not, and identifies collocates based on concordance reading, Stubbs means only habitual

co-occurrence, and identifies collocates using statistical measures. Partington relies on individual co-occurring words, which he calls “collocates” on some occasions and “items” on other occasions, as well as contextual expressions of evaluation in his identification of semantic prosody (see section 2.5.2.1). In this thesis, Stubbs’ approach will be employed. Only statistical collocates will be taken into account when identifying the semantic prosody of a word.

As we saw earlier, Louw, Stubbs, and Partington generally restrict semantic prosody to being only positive or negative (see section 2.5.2.2), although Stubbs, in his early work, also allows semantic prosody to be mixed or neutral. In particular, Stubbs (1995: 16) argues that not every lemma has as clear a semantic prosody as the lemma *cause*, which has a clear negative semantic prosody (see section 2.3.3.1). Examining the lemma *create*, for instance, Stubbs (1995: 16) asserts that the lemma is “prosodically mixed or incomplete”, because there is a mixture of positive, negative and neutral collocates. Stubbs also allows semantic prosody to be neutral. He argues that the lemma *reason* has a “largely neutral” prosody (Stubbs 1995: 16). However, in this case, he does not explicitly state his reasoning or what he means by a neutral prosody. It appears from the examples of the collocates of *reason* provided by Stubbs that the lemma is argued to have a neutral prosody because the majority of its collocates are semantically neutral.

Unlike Stubbs, Louw and Partington seem to allow semantic prosodies only to be positive or negative, even though Partington (2004: 144) also argues in some cases that a word does not incline towards having either a positive or negative semantic prosody (which is, in effect, a neutral prosody).

One methodological issue that arises here, but which seems to be left unstated in the literature, is exactly how large a proportion of positive or negative collocates is required for a word to be said to have a positive or negative semantic prosody. Louw, Stubbs, and Partington do not address this point explicitly, although I would argue that each scholar has his own rule of thumb.

For instance, Louw (1993: 160) assigns a negative semantic prosody to *utterly* based on the evidence that “there are few ‘good’ right-collocates”. When we look at his concordance lines for *utterly*, we see that the negative collocates outnumber the positive ones. Louw’s rule of thumb thus seems to be identifying the semantic prosody according to the majority of collocates. If a word has more positive than negative collocates, it will be argued to have a positive semantic prosody, and vice versa. However, Louw does not indicate how great the proportional difference between positive and negative collocates must be for a word to be considered to have a particular semantic prosody. In addition, he does not consider or even refer to neutral collocates.

Stubbs seems to adhere to the same rule as Louw. However, as previously discussed, in cases where there is a mixture of positive, negative, and neutral collocates and in cases where the majority of collocates are neutral, Stubbs speaks in terms of a mixed prosody and a neutral prosody, respectively.

In his analysis of the *happen* word group, Partington (2004: 135-144) explicitly classifies what he calls the “environments” into three meaning categories: positive, negative, and neutral (see section 2.5.2.1). That said, he does not seem to take the neutral instances into account when identifying semantic prosody. For example, Partington (2004: 137) analyses the lemma *happen*. He reports that of 150 instances, 76 occurs in

neutral environments, 54 in negative environments, and 20 in positive environments. Even though the preponderance of the instances of *happen* occurs in neutral environments, he does not assign a neutral semantic prosody to *happen*. He rather argues that *happen* has a negative semantic prosody. A similar case is the lemma *occur*, to which Partington (2004: 141, 144) assigns a negative semantic prosody, despite the fact that most of the instances of the lemma are found in neutral environments. Of the lemma's 150 instances, 100 occur in neutral environments, 36 in negative environments, and 14 in positive environments. Given these examples, I would argue that Partington follows a rule of thumb of ignoring instances where the lemma occurs in neutral environments. What he takes into account is whether there are more positive or negative instances. If negative instances outnumber positive instances, as seen with the cases of *happen* and *occur*, even if there are yet more neutral instances, a negative semantic prosody will be assigned (and vice versa).

Even though both *happen* and *occur* are argued to have a negative semantic prosody, they do not have the same prosodic strength. Partington argues that *happen* has a stronger negative prosody than *occur* (see section 2.3.4.1). Partington does not spell out his reasoning for this. However, from the numbers Partington provides, it seems that he bases his judgement on the number of negative instances. Since *happen* has more negative instances (54) than *occur*, it is said to have a stronger negative semantic prosody than *occur* (36). Likewise, if we consider the ratio between the negative instances and the positive instances of *happen* and *occur*, we see that it is greater for *happen* (2.7 times more negative instances) than for *occur* (2.5 times), although not by very much.

In the cases of *happen* and *occur*, the difference in number between negative and positive instances is quite big. There are, as we have seen, more than twice as many negative as positive instances. However, Partington does not explicitly discuss any cases where there is only small difference between the number of positive and negative instances, or cases where almost all instances are neutral. This is one factor making it difficult to work out the rule of thumb he uses.

We have seen that even though the three scholars generally approach semantic prosody in the same way, there are some points of differences in their methods and in their unstated criteria for what makes a semantic prosody positive or negative. In this thesis, as pointed out earlier, I will employ Stubbs' approach of identifying the semantic prosody of a word based on statistical collocates. It has been mentioned that in his analysis of the *happen* word group, Partington classifies what he calls "environments" into three meaning categories: positive, negative, and neutral. In this study, I will adopt Partington's practice and likewise classify collocates into positive, negative, and neutral. I will consider whether there are more positive collocates or more negative collocates. Only when the difference in the proportion between positive collocates and negative collocates is at least threefold will I argue that a word has a clear positive or negative semantic prosody. In cases where the difference is less than threefold, I will argue that the word does not have either a positive or a negative semantic prosody. However, if 70% or more of the collocates are neutral, the word will be argued to not to have any clear positive or negative prosody, even if the difference in the proportion between positive and negative collocates is threefold or more. In such cases with no clear positive or negative semantic prosody, I will refer to the word as having a neutral semantic prosody.

A discussion of the number of the words whose semantic prosody will be examined as well as the criteria for word selection, will be separately provided in each analysis chapter (for Chapters 4, 5 and 6 in sections 4.3, 5.4 and 6.4 respectively). The statistical measure of collocational strength to be used in the analysis and the criteria for collocate selection, including a cut-off point and a span, will be discussed in section 4.2.1.

3.4.1.2 Type vs. token frequency

Whether to count token frequency or type frequency of the collocates is another methodological issue that is not explicitly addressed in the literature. Token frequency is the total number of instances of the word under consideration. Type frequency, on the other hand, is the number of different wordforms (see McEnery and Hardie 2012: 252 for the definition of *token* vs. *type* in corpus linguistics). Although Louw, Stubbs, and Partington do not address this issue explicitly, their approach can be worked out from their discussion of semantic prosody, as follows.

When Louw examines the concordance lines of *utterly*, he manually identifies the collocates of *utterly* in each line. Based on the dominance of the negative collocates identified, Louw argues that *utterly* has a negative semantic prosody. That Louw assigns the negative prosody by considering each individual collocate instance suggests that he counts token frequency. Each individual instance of a collocate is counted, and as most of these instances of collocates are negative, *utterly* is argued to have a negative semantic prosody.

Partington also appears to count token frequency. He states explicitly that he examines concordance lines in order to identify the semantic prosody of a lemma. For

example, Partington (2004: 135) states “Looking at the concordance lines of these items, there can be no doubt that SET IN has an extremely unfavourable prosody.” Therefore, as with Louw, Partington considers each individual instance of a co-occurring item. In the case of *set in*, the lemma is argued to have a negative prosody because most of the instances are of negative items.

Stubbs, on the other hand, appears to count type frequency. Instead of examining a concordance and manually extracting collocates in each line, Stubbs typically generates statistical collocates and identifies semantic prosody based on the collocates. Under this approach, only statistically significant collocates are counted as contributing to the semantic prosody of a word. For example, Stubbs (1995: 17) assigns a negative semantic prosody to the lemma *affect* on the basis that most of its significant collocate types are negative. This counts each collocate wordform once only. However, Stubbs also uses token frequency in some of his research by prioritising frequent collocates. For instance, examining the semantic prosody of the lemma *undergo*, Stubbs identifies the 20 most frequent collocates from the lemma’s concordance and argues that *undergo* has the discourse prosodies of “involuntary” and “unpleasant” (see section 2.3.3.2).

In sum, then, Louw and Partington count token frequency, whereas Stubbs sometimes counts type frequency and sometimes counts token frequency. It is arguably reasonable to count token frequency when identifying semantic prosody. Semantic prosody is a collocational phenomenon where a word shows a tendency to repeatedly co-occur with positive or negative words. Partington argues for this point in terms of Hoey’s lexical priming theory (see section 2.3.4.1). Hoey contends that a word is primed for collocational use. Each word carries with it information about its semantic prosody, that

is, whether it tends to occur in positive or negative contexts. This information is established through the language user's repeated encounters with the word. These repeated encounters strengthen the user's mental connections between the word and its semantic prosody. On these grounds, it would be justifiable to count every individual token of every collocate type when establishing positive vs. negative semantic prosody.

However, from a psychological perspective, type frequency is *also* important. For instance, Bybee and Hopper (2001 cited in Ellis 2012: 9) argue that the more *different* words can appear in a particular position in a syntactic construction, the less likely it is that a construction is related to a specific word. Rather, when many different types appear in a given position, a general pattern is likely to be created, and as the pattern is generalised, it is likely to extend to new words. In the case of semantic prosody, it would follow that the more *different* negative words a word co-occurs with, the more likely the word is to co-occur with *other* negative words (and the same would apply to positive words). We can thus say that such a word displays a stronger negative semantic prosody than one which has fewer negative collocate types.

Thus we see that both token frequency and type frequency play a role in a word's semantic prosody. In this thesis, I will count type frequency for the reasons discussed above. The selection of type frequency over token frequency is also a practicality issue. Because the corpus used in the thesis is quite big, it would be difficult in practice to analyse every single co-occurring token. Moreover, because statistical measures extract frequent collocate types and ignore individually infrequent types in the first place, token frequency is in effect automatically taken into account when working from a statistical collocate list. With the English data, I will use type and token counts based on lemmas

rather than wordforms. However, this is not necessary when working with the Thai data. Unlike English, Thai lacks inflections, and each Thai word thus has one grammatical form (see section 1.3.3.2).

There remains, however, one way in which token frequency may be important: within a list of collocate types meeting some statistical criteria, some types may represent many more tokens than others (and thus, a greater fraction of the instances of the word under study). This does in fact happen in my data. This could potentially skew the analysis if the low-frequency collocate types show opposite evaluative behaviour to one or more high-frequency types. To ensure that this did not happen to my data, I examined six pairs of translation-equivalents (data from my analysis in Chapter 6) that display the same positive or negative semantic prosody, but this time I also counted token frequency of positive and negative collocates. In all these six instances, taking into account token frequency does not change the semantic prosody of the translation pairs. Thus, I am satisfied that utilising an approach based on the count of type frequency is methodologically valid.

3.4.1.3 Inter-rater reliability

The classification of collocates as positive, negative, or neutral is another methodological issue that should be considered. To the best of my knowledge, researchers that adopt the polarity-oriented approach in their studies of semantic prosody, such as Louw (1993), Stubbs (1996), Partington (1998; 2004; 2014), and Xiao and McEnery (2006), use their introspection as a native speaker of the language(s) under study to classify collocates. This is the method I wish to employ in my study. However, it

might be possible to object to this approach, on the basis that my introspection could be misguided, in the sense of not adequately reflecting the introspections of native speakers of Thai in general. To address this objection, I conducted a small-scale study of inter-rater agreement in order to justify the use of my introspection as a native speaker of Thai to classify Thai collocates.

In this study, a questionnaire consisting of a list of forty Thai words was given to ten competent native speakers of Thai. Eight of these participants are PhD students, and two hold a master's degree. These forty Thai words were chosen from the collocate lists of the nodes /kreençay/ 'considerate', /kòðhâykàət/ 'cause', and /chôp/ 'like', the objects of the study of Chapter 4. All 40 had been classified as positive, negative, or neutral by me, but in some cases I was sure of the classification, and in other cases I was not. The participants were asked to classify the words as positive, negative, or neutral. They were also allowed to say 'not sure'. The questionnaire is given in Appendix 2.

The results show that the participants are generally in agreement. Table 3.4 presents the level of agreement among the participants.

Number of participants selecting the most popular option	Number of words	Percentage	Cumulative percentage
10	13	33	33
9	7	18	51
8	10	25	76
7	3	7	83
6	3	7	90
5 or less	4	10	100

Table 3.4 Level of agreement among participants

Table 3.4 shows that out of 40 cases, there are 36 with consensus among the participants. In fact, there are 13 cases where the participants are in perfect agreement. There are only four cases where there is no consensus. These four cases are /khlûanwăy/

‘move’, /sǒmdun/ ‘balance’, /phàatphǒon/ ‘adventurously’, and /prīaw/ ‘sour’. From my perspective, /khlûanwǎy/ ‘move’, /phàatphǒon/ ‘adventurously’, and /prīaw/ ‘sour’ are hard to decide, because their evaluative meaning sometimes depends on context in which they occur. Finally, /sǒmdun/ ‘balance’ is in my view slightly positive, which would explain why the participants were split between positive and neutral.

Out of 40 cases, there are 37 cases (92.5%), including the four cases where there is no consensus, where my judgement is in agreement with that of the majority of the participants. These 37 include words I was not very sure of, such as /kèptua/ ‘introvert’ and /khwaamkhûanwǎy/ ‘movement’. I had classified /kèptua/ ‘introvert’ as negative and /khwaamkhûanwǎy/ ‘movement’ as neutral. My judgement is in agreement with that of the majority of the participants. Eight participants classified /kèptua/ ‘introvert’ as negative, and seven classified /khwaamkhûanwǎy/ ‘movement’ as neutral.

The three cases where disagreements occur are /phùukphan/ ‘bond’, /piinpàay/ ‘climb’, and /sǎŋsǎn/ ‘socialise’. The majority of the participants classified /phùukphan/ ‘bond’ as positive. I classified the word as neutral, as in addition to the most frequent context where the word is used to refer to a bonding between people, which is positive, the word is used in legal contexts, where it refers to an obligation, which is negative. I thus classified the word as neutral. /piinpàay/ ‘climb’, for me, is neutral. The majority of the participants classified the word as negative. Finally, I classified /sǎŋsǎn/ ‘socialise’ as neutral, but most of the participants classified the word as positive. I had actually classified the word as positive at first, but then changed my mind.

It can be seen that, of the three cases of disagreement, there is no case where I classified the word in the opposite direction from the majority of the participants. That is,

there is no case where the majority of the participants classified the word as positive, and I classified it as negative, or vice versa. Thus, if my judgement does differ from that of other speakers at some point, the effect of this would most likely only weaken the strength of the semantic prosody. It is unlikely to affect the quantitative result of the analysis by pushing the semantic prosody in another direction.

In sum, this small-scale study of inter-rater agreement has thus shown that first, the participants are generally in agreement, and second, my introspection is generally reflective of the introspection of other native speakers of Thai. For this reason, I am satisfied that the use of my introspection in classifying the Thai collocates is methodologically valid. That said, we have seen not only cases where the participants are in perfect agreement, but also cases where disagreement emerges, including some cases where my classification is different from that of the majority of the participants. Therefore, in my analysis in Chapters 4 and 6, there might likewise be some cases where other analysts would disagree with my classification and tend to classify the collocates differently from me. This is regrettable, but not avoidable.

It is also worth noting that in my classification of the collocates as positive, negative, or neutral, I consider their meaning in isolation rather than their meaning in context. That is, I look at the collocates, consider their meaning on their own, and classify them according to my introspection. So I consider, for example, *good* as positive and *bad* as negative without taking into account what is being labelled good or bad in the actual text. These two examples are easy cases, because what is good is good, and what is bad is bad. However, there are also difficult cases that might cause controversy, especially collocates that denote movements, such as *increase* and *decrease*. It might be argued that

the evaluative meaning of these collocates is generally dependent upon the entity described, that is, what it is that increases or decreases. For example, while “salary increases” is positive, “pressure increases” is negative. Nevertheless, in this case and other similar cases, I still consider the meaning of the collocate in isolation. Looking at the context where an *increase* occurs would be straying away from the definition and method established for the polarity-oriented approach – that is, identifying a node’s evaluative potential based on its positive or negative collocates (see section 3.4.1). Moreover, a problem of looking at the context of the collocates of *increase* is that it opens the prospect of an infinite regress of analysis. That is, it would mean that I will have to identify, to a certain degree, the semantic prosody of each collocate of *increase* via its own collocates before I could identify the semantic prosody of the node; and that, in turn, might require the hidden evaluative potential of each of those second-order collocates to be identified; and so on *ad infinitum* with no principled way to determine an end point. Given these two points, I adopt the practice of considering only the meaning in isolation of the direct collocates of the node throughout the analysis, even in cases where the evaluative meaning of the collocate may appear to depend on that collocate’s own further co-occurrence behaviour.

Let us now turn to the classification of the evaluative meaning of collocates that denote movements. It may seem difficult to assign evaluation to *up* or *down* in isolation. However, it has frequently been observed that in their metaphorical senses, at least, these words do in fact tend to have a consistent evaluation in themselves. For instance, Lakoff and Johnson’s (1980: 14-16) classic study of conceptual metaphors considers the orientational metaphors *up* and *down*, arguing that the concept of *up* is associated with

good, whereas the concept of *down* is associated with *bad*. For the purposes of a collocate analysis, then, *increase* and other words that denote the idea of something going up can arguably be classified as good, whereas *decrease* and other words that express the idea of something going down could arguably be classified as bad.

3.4.2 Sinclair's approach (the EUM-oriented approach)

Within the Sinclairian approach, semantic prosody is viewed as a pragmatic meaning that is spread across an extended unit of meaning (see section 2.3.2.1). It is a compulsory element of an extended unit of meaning. The method adopted for identifying semantic prosody according to this approach is thus to observe the concordance lines of the word under study.

For each word, 200 concordance lines will be randomly selected. 200 is large enough to reveal common patterns around the word; samples of around this size, or sometimes less, are commonly used in the literature. Partington (2004: 133-144), for example, looks at 150 concordance lines in his analysis of the *happen* word group (see section 2.3.4.1). Baker (2006: 74) relies on 140 concordance lines in his investigation of the terms *refugees* (see section 2.6). Wei and Li (2013: 112-133) examine 100 concordance lines in their analysis of the semantic prosodies of English-Chinese translation pairs (see section 2.7.5). Tognini-Bonelli (2009: 209) looks at only 49 concordance lines in her analysis of the semantic prosody of *andare incontro* (see section 2.7.7). Analysing more than 200 concordance lines, say 400, might reveal more patterns, but would be much more time-consuming, and could make the analysis impractical, especially when it is necessary to analyse a large number of words, as will be the case in

Chapter 5. The patterns around the word will be identified according to Sinclair's model of the extended unit of meaning in terms of colligation, collocation, semantic preference, and semantic prosody. Within this approach, semantic prosody is not restricted to being just only positive or negative, but can be any pragmatic function or meaning.

For a pattern around the word under study to be classified as an extended unit of meaning, it should meet two criteria derived from Sinclair's analysis as reviewed in the previous chapter. First, the pattern must have a fixed unit as its core. The core can be formed by the word under study alone or by the word under study plus one or more other fixed words adjacent to it. Second, the pattern must have a clear pragmatic function that is distinct from a combination of the meanings in isolation, either literal or metaphorical, of the individual parts of the pattern. Moreover, this pragmatic meaning must be spread across the whole pattern, rather than being evident from any individual element of the pattern. This last is the most important compulsory criterion for a pattern to be considered as an extended unit of meaning.

A further element that I will look for in a pattern is specific colligations and semantic preferences. Colligations and semantic preferences are often, but not always necessarily, present in an extended unit of meaning. In cases where they exist, they must belong the whole pattern and not to any individual word in it. However, the absence of colligations and semantic preferences will not be taken to prevent something from being considered as an extended unit of meaning.

There may occur cases where an identified pattern seems to be an extended unit of meaning, but there is not enough evidence for it. For example, we might observe a pattern which meets the criteria specified above, but there are not enough examples to argue

confidently for that pattern's status as an extended unit of meaning. In such a case, I will not say that the pattern is an extended unit of meaning, but rather will note it as something which might turn out to be a unit if further evidence could be obtained.

Throughout my analysis, I will be *conservative*, in the sense that I will be strict about the application of the criteria specified above. I will state that the pattern identified is an extended unit of meaning in the Sinclairian sense only when I am confident that the pattern meets the two criteria specified. In case where I am not completely confident that the pattern meets those criteria, I will not state that it is an extended unit of meaning. Applying this conservative standard means that I may not be able to claim to have discovered as many extended units of meaning as could be derived by a researcher applying a less strict standard. However, it also means that I will only argue for units that I can be very confident of, and thus avoid the risk of arguing for spurious units in my results.

3.5 Concluding remarks

In this chapter, I presented a brief overview of my data: the Thai National Corpus and the British National Corpus. I also detailed the methodology that I will adopt in my analysis. This involved a discussion of Louw, Stubbs, and Partington's approach, which I also refer to as *the polarity-oriented approach*, and a discussion of Sinclair's approach, which I also refer to as *the EUM-oriented approach*. In particular, in my discussion of the polarity-oriented approach, I have established the criteria for what makes a semantic prosody positive or negative, which has previously not been entirely clearly stated in the literature. I have also argued for using a count of type frequency rather than token

frequency in the polarity-oriented approach, and supported the validity of the use of my introspection as a native speaker of Thai for the classification of collocates by conducting a small-scale study of inter-rater agreement. In the next chapter, I will investigate the advantages and disadvantages of the two primary approaches to the study of semantic prosody in Thai, using the methodologies established in this chapter.

Chapter 4 – Comparison of the polarity-oriented approach and the EUM-oriented approach

4.1 Introduction

The aim of this chapter is to address the first research question: what are the advantages and disadvantages of the major approaches to semantic prosody proposed in the literature for describing semantic prosody in Thai? In section 4.2, I present the specific approaches adopted in the analysis. Then I introduce the words whose semantic prosody will be analysed in section 4.3. In section 4.4, I present the results of the analysis, with a detailed discussion of the results in section 4.4.1.4 and section 4.4.2.4.

4.2 Approaches to the analysis

Since the aim of the first research question is to investigate the advantages and disadvantages for describing semantic prosody in Thai of the major approaches to semantic prosody proposed in the literature, three selected words, which will be discussed in section 4.3, were examined by two different methods, as follows.

4.2.1 Louw, Stubbs, and Partington's approach (the polarity-oriented approach)

As outlined in section 3.4.1.1, Louw and Stubbs heavily rely on collocate analysis in their identification of semantic prosody; Partington looks at collocates as well as longer stretches of text. For this part of the analysis, I mainly looked at the statistical collocates of the selected words. I chose to look at the collocates within a 4-4 window span as suggested by Sinclair *et al.* (2004: 5). The statistical measure of collocational strength used was Log Ratio (Hardie forthcoming). Log Ratio, Hardie argues, is a better

statistic for keywords than log-likelihood. It is an “effect-size” statistic and “represents how big the difference between two corpora is for a particular keyword” (Hardie forthcoming). The Log Ratio statistic can also be used for collocation. In this case, the Log Ratio score of a collocate represents how much more frequent the collocate is near the node than elsewhere. For example, a collocate with a Log Ratio score of 1 occurs near the node two times more frequently than it occurs elsewhere, and “every extra point in Log Ratio score represents a doubling in size of the difference between the collocate’s frequency near the node and its frequency elsewhere” (Hardie forthcoming). For this analysis, only items with a Log Ratio score of 3 or more that occur in at least five different texts were considered as collocates of a given node. That is, only items that occur at least eight times more frequently near the node than elsewhere and occur in at least five different texts were considered collocates. I chose a Log Ratio score of 3 as a cut-off point, because Log Ratio is very similar to the Mutual Information statistic, and Hunston (2002: 71) suggests that items with an MI-score of three or more can be considered to be significant.

Under this approach, semantic prosody is restricted to being positive or negative (or neutral) following the rule of thumb outlined in section 3.4.1.1. In section 4.4, I present the lists of collocates in the format that Stubbs (1995: 15-18) uses. The lists are sorted by Log Ratio, but the actual Log Ratio scores are not included, since Stubbs likewise does not present statistical scores in his lists of collocates. I extend Stubbs’ format in one way, namely, by including highlighting: I will underline collocates with a positive meaning, and present collocates with a negative meaning in bold. Collocates with a neutral meaning will be left unhighlighted.

4.2.2 Sinclair's approach (the EUM-oriented approach)

To investigate Sinclair's approach to semantic prosody, I examined 200 randomly-selected concordance lines for each of the selected words (see section 3.4.2). I identified the major patterns around these words according to Sinclair's model of the extended unit of meaning, looking for colligation, collocation, semantic preference, and semantic prosody (see section 2.3.2.1 and section 3.4.2). Under this approach, a semantic prosody can be any pragmatic function or meaning. I will present each proposed extended unit of meaning in a one-line format, using the notations in Table 4.1.

Notation	Meaning
Bold font	Lexical core of an extended unit of meaning.
(...)	Items in round brackets are optional.
[...]	Square brackets contain an explanation of what is found in this position (grammatical/semantic restriction).
Column	Items in a column are alternatives in the given position.
	Sentence break.

Table 4.1 Notations for extended units of meaning

Using these notations, for example, Sinclair's extended unit of meaning whose core is *naked eye* (see section 2.3.2.1) would be represented as follows:

[some pragmatic expression of difficulty] [verb/adjective expressing visibility] with the **naked eye** to

So in the above example, *naked eye*, which is presented in bold, is the wholly fixed lexical core of the unit. The article *the* is rather fixed; it almost always appears right before *naked eye*. The prepositions *with* and *to* are also partly fixed, but they are presented in a column to indicate alteration. That is, the element in this position must be either *with* or *to*. At the next two positions leftwards, pairs of square brackets are used,

containing short prose characterisations of what is required in these two positions. So the bracketed element [verb/adjective expressing visibility] means that what is required in this position is a verb or adjective that expresses visibility, and the bracketed element [some pragmatic expression of difficulty] means that what is required in this position is a word or phrase that expresses the pragmatic expression of difficulty. These elements are, like the fixed elements, compulsory; otherwise, they would have been put in round brackets.

4.3 Selection of Thai words

In this analysis, I will have to investigate each word using two approaches; for that reason, it follows that I cannot look at more than a small number of words. On the other hand, if I investigate only one or two words, we might suspect that the results obtained relate only to these specific words, rather than being generalisable. I have therefore opted to look at three words. While it would have been equally possible to look at four or five, this would have made the analysis much more extensive for minimal further benefit in terms of comparing the approaches. The three words selected for this analysis are /kreeŋcay/ ‘considerate’, /kòŋhâykàt/ ‘cause’, and /chôp/ ‘like’ (in the sense of enjoying (doing) something). Each of these words was selected on the basis of a different motivation. The word /kreeŋcay/ is interesting to me because there seems to be no word in English that has exactly the same meaning as /kreeŋcay/. Although it is a verb in Thai, most of the possible English translations are adjectives. The closest translation-equivalent is probably ‘considerate’ or ‘reluctant,’ as in ‘reluctant to impose on a person’. /kòŋhâykàt/ is a translation-equivalent of English *cause*, which has been established to

display a negative semantic prosody (Stubbs 1995: 3). It will thus be interesting to see if /kòhâykàət/ also has a negative semantic prosody. Finally, my personal impression is that /chôp/ is normally used in a negative context, and I would like to see if the word really does have a tendency to occur in unfavourable environments.

4.4 Results

4.4.1 Louw, Stubbs, and Partington's approach

4.4.1.1 /kreeŋcay/ 'considerate'

There are 1,009 instances of /kreeŋcay/ in the corpus. The collocates of the word include (in order of Log Ratio score):

/khii/ (grammatical particle), /kreeŋ/ 'fear', /róopkuan/ 'bother', /caŋ/ 'quite', /khawróop/ 'respect', /klâa/ 'brave', /klua/ 'afraid', /náamsiŋ/ 'tone', /penray/ (part of /mây penray/ 'Don't worry'), /tâataŋ/ 'gesture', /ríusək/ 'feel', /ròk/ (pragmatic particle), /kənpay/ 'too much', /ʔəy/ 'utter', /phômêe/ 'parents', /phûuyà/ 'adults'

Of the 16 collocates of /kreeŋcay/ that meet the criteria given in section 4.2.1, 5 appear to have a negative meaning. There are two positive collocates and nine neutral collocates. It might thus be concluded that /kreeŋcay/ does not have either a positive or a negative semantic prosody. That is, it shows a neutral semantic prosody.

4.4.1.2 /kòhâykàət/ 'cause'

There are 4,685 instances of /kòhâykàət/ in the corpus. The collocates of the word include (ranked by Log Ratio score):

/rákhaaykhuəŋ/ ‘irritate’ */phõnsĩa/* ‘negative effect’ */khwaamsĩahǎay/*
 ‘damage’ */pháyantàraay/* ‘danger’ */monphaawá/* ‘contamination’
/phõnráay/ ‘negative effect’ */dùatròɔn/* ‘in trouble’ */monphít/* ‘pollution’
/pànpùan/ ‘frantic’ */tèekyêek/* ‘disunited’ */yàylǔəŋ/* ‘enormously’
/sàthuancay/ ‘emotionally hurt’ */phõnkràtóp/* ‘effect’ */təŋkhríat/*
 ‘serious’ */lamʔiaŋ/* ‘bias’ */ʔantàraay/* ‘danger’ */wítòkkaŋwon/* ‘worried’
/ramkhaan/ ‘annoyed’ */khlúanwǎy/* ‘move’ */phõnláp/* ‘result’
/sùamsoom/ ‘deteriorate’ */sùamsĩa/* ‘tarnished’ */yûŋyâak/* ‘complicated’
/sàpsõn/ ‘confused’ */sĩahǎay/* ‘damaged’ */ráayrɛŋ/* ‘serious’
/khwaamkhríat/ ‘stress’ */câaŋŋaan/* ‘employ’ */khûusǎnyaa/* ‘party’ */lùam/*
 ‘unequal’ */khwaamkhàtyéŋ/* ‘conflict’ */ròokmáreŋ/* ‘cancer’ */lúklaam/*
 ‘spread (of disease)’ */wûnwaay/* ‘in confusion’ */máreŋ/* ‘cancer’ */núi/*
 ‘debt’ */wípâakwícaan/* ‘criticise’ */rɛŋkòtdan/* ‘pressure’ */sìŋwêetlóm/*
 ‘environment’ */wítòk/* ‘worried’ */khwaamkhlúanwǎy/* ‘movement’
/ʔànaamay/ ‘hygiene’ */sĩapriap/* ‘disadvantageous’ */phlêətphləən/* ‘enjoy’
/plianplɛəŋ/ ‘change’ */kèe/* ‘for’ */sǎnti/* ‘peace’ */sàthĩanráphâap/* ‘stability’
/phõn/ ‘result’ */panhǎa/* ‘problem’ */ʔàatyaaakam/* ‘crime’ */pràyoət/*
 ‘benefit’ */wêetlóm/* ‘surround’ */phanthúkam/* ‘heredity’ */dâypriap/*
 ‘advantageous’ */sàmǎəphâak/* ‘equal’ */thòkthĩaŋ/* ‘dispute’ */nítikam/*
 ‘juristic act’ */sǎmphantháphâap/* ‘relationship’ */yóəm/* ‘naturally’
/wâaŋŋaan/ ‘unemployed’ */khàatkhlɛən/* ‘lack’ */sũunsĩa/* ‘lose’ */raaydây/*
 ‘income’ */pàtikiriyaa/* ‘reaction’ */ròpkuan/* ‘bother’ */sòmduŋ/* ‘balance’
/ʔuppàsàk/ ‘obstacle’ */phùukphan/* ‘bond’ */pràthêetchâat/* ‘nation’
/khwaamklua/ ‘fear’

It is clear from the predominance of negative collocates that */kòəhâykəət/* has a tendency to occur in semantically negative contexts. Of the 71 collocates, 44 have a negative meaning. There also exist neutral and positive collocates. Further investigation into these collocates reveal that some of them are in fact frequently used in negative contexts. */sìŋwêetlóm/* ‘environment’, for example, does not in itself indicate any affective meaning. However, in many of the 78 instances, it is used in negative environments. One obvious instance is its use as part of the noun phrase */panhǎa sìŋwêetlóm/* ‘environmental problem’, which occurs 11 times. It also appears in the noun phrase */phõnsĩa tòə sìŋwêetlóm/* ‘negative effect on the environment’ three times. A similar example from the positive collocate list is */sàthĩanráphâap/* ‘stability’. Of the 13 instances of */sàthĩanráphâap/* as a collocate of */kòəhâykəət/*, 6 appear as part of */khwaam*

mây mii sàtīanráphâap/, /khwaam ráy sàtīanráphâap/, or /kaan ráy sàtīanráphâap/, all of which mean ‘instability’, and one appears as part of the noun phrase /panhãa sàtīanráphâap/ ‘stability problem’. However, there are also cases where the neutral and positive collocates actually do reflect neutral and positive environments respectively. For instance, /pràyòot/ ‘benefit’ occurs 174 times as a collocate of /kòhâykàət/, mostly immediately after /kòhâykàət/, and functioning as a direct object (i.e. ‘cause a benefit’).

It can thus be concluded that like English *cause*, /kòhâykàət/ has an overall negative semantic prosody. However, the fact that it has a strong tendency towards co-occurrence with negatively evaluated items does not prevent it from appearing in neutral or even positive environments, albeit less often.

4.4.1.3 /chôp/ ‘like’

/chôp/ appears 20,942 times in the corpus. Its collocates ranked by Log Ratio score include:

/chôp/ ‘like (emphatic reduplication)’ /sămmaasàtì/ ‘mindfulness’ /sămmaasàmaathí/ ‘right concentration’ /**khîn**âa/ ‘face’ /taamcay/ (part of /taam cay chôp/ ‘as you please’) /**sò**trúusòthên/ ‘snoop’ /**sũ**ñsĩŋ/ ‘keep company with’ /phaatphõn/ ‘adventurous’ /trèe/ ‘hang around’ /chít/ (person’s name) /**wa**ñthâa/ ‘act big’ /**cù**kcik/ ‘fussy’ /**kh**lùk/ ‘absorbed in’ /**cha**ŋ/ ‘hate’ /phet/ ‘spicy’ /pràphrút/ ‘behave’ /**pũ**anpīan/ ‘loiter’ /cíap/ (person’s name) /yèè/ ‘tease’ /damrì/ ‘think’ /**thú**tcàrit/ ‘corrupt’ /yĩŋ/ ‘woman’ /**tũ**u/ ‘pester’ /heehaa/ ‘enjoy oneself’ /**cũ**ucĩi/ ‘fussy’ /**chó**ktòy/ ‘have a fight’ /laaŋ/ (part of an idiom ‘one man’s meat is another man’s poison’) /piinpàay/ ‘climb’ /**sũ**kson/ ‘naughty’ /líañchĩp/ ‘make a living’ /búu/ ‘action’ /nísây/ ‘characteristic’ /**?ò**ò?uàt/ ‘boast’ /**ke**eree/ ‘mischievous’ /mí/ ‘no’ /?ùppànísây/ ‘characteristic’ /klaañcêeŋ/ ‘outdoors’ /tèeŋtua/ ‘get dressed’ /khúkkĩi/ ‘cookie’ /tũmhũu/ ‘earring’ /?àdirèek/ ‘hobby’ /**pr**àcòp/ ‘flatter’ /**kè**ptua/ ‘introvert’ /núuphleey/ ‘lyric’ /yúy/ (person’s name) /**h**ãarũuŋ/ ‘pick a quarrel with’ /thamtua/ ‘act’ /**ph**lèeŋ/ ‘peculiar’ /**rũ**uyphũy/ ‘aimlessly’ /lâap/ (Thai dish) /**?aw**priap/ ‘take advantage of’ /mâykhòy/ ‘not quite’ /**c**ãwchũu/ ‘flirty’ /**ra**ŋkæ/ ‘bully’ /phàconphay/ ‘take adventure’ /**nin**thaa/ ‘gossip’

/khwaamchôp/ ‘liking’ /chôpcay/ ‘pleased’ /khôŋwăan/ ‘dessert’ /sôn/ ‘heel’ /prîaw/ ‘sour’ /chôp/ ‘shop’ /săŋsăŋ/ ‘socialise’ /sàtaay/ ‘style’

Of the 64 collocates, 26 have a negative meaning. It is interesting that many of these negative collocates are verbs. Thus, it might be said that /chôp/ has a negative semantic prosody, which is especially strongly expressed by verbal collocates.

4.4.1.4 Discussion

The above results fit with my expectations in the cases of /kòhâykəət/ ‘cause’ and /chôp/ ‘like’. The semantic prosody of *cause* has been extensively studied in the literature. It has been established to display a negative semantic prosody in English (Stubbs 1995: 3). Its translation equivalents in Danish and Chinese have also been found to display a negative semantic prosody (Dam-Jensen and Zethsen 2008: 1618; McEnery and Xiao 2006: 117). Therefore, the fact that /kòhâykəət/ also has a negative semantic prosody does not come as a surprise.

Before investigating the corpus, I had the impression that /chôp/ is normally used with a negative verb, such as /máw/ and /ninthaa/, which mean ‘gossip’, in a serial verb construction to indicate negatively-evaluated personal habits. The word /ninthaa/ does appear as a collocate of /chôp/. /máw/ in fact also appears in the collocate list, but as it appears in only three different texts, I did not consider it in my analysis. Of the 19 total instances of /ninthaa/, 14 appear after /chôp/ in a serial verb structure such as /chôp ninthaa/, /chôp maa ninthaa/, or /chôp phùut ninthaa/, all of which mean ‘like to gossip’.

There are three interesting points about /chôp/. First, intuitively, /chôp/ seems to be a positive word in terms of its core semantics and connotations and thus should occur

in favourable environments (people generally “like” doing positive things). Therefore, its occurrence in unpleasant environments seems to contradict our intuition. Second, the fact that the negative semantic prosody of /chôp/ seems to be linked with a particular colligational set (verbs), especially in the serial verb structure, supports Hunston’s argument (in support of Sinclair) that semantic prosody is a property of a unit of meaning, rather than a property of a word (Hunston 2007: 258). Therefore, it might be contended that within this serial verb construction, /chôp/ is the core of a unit of meaning of which the colligation with a verb is another component. These verbs in turn express the semantic preference for negativity so that, overall, the unit can be said to have a negative semantic prosody, specifically referring to a personal bad habit (see further section 4.4.2.3). Third, English *like*, the translation-equivalent of /chôp/, does not display the same negative prosody as its Thai counterpart. Examining the BNC, I found that most verbal collocates of *like* have either a positive or a neutral meaning.

Regarding /kreeŋcay/, my only personal impression about this word is that it tends to be used in contexts that express that /kreeŋcay/ (“being considerate”) is a quality that (Thai) people are expected to have. I did not have any intuition regarding its semantic prosody. In line with this, the collocates of /kreeŋcay/ do not seem to indicate a positive or negative semantic prosody. Employing Sinclair’s approach may, however, yield a different result (see section 4.4.2).

The analyses show that simply looking at single-word collocates may not be sufficient to assess semantic prosody. We have seen this disadvantage in the case of /kôhâykàt/ where, for example, the neutral word /sìŋwêtlóm/ ‘environment’ is actually part of the negative noun phrase /panhãa sìŋwêtlóm/ ‘environmental problem’.

Fortunately, since the negative collocates still dominate the collocate list, the practice of looking at the collocates alone does not skew the results in this case.

There appear to be some problems caused by the corpus itself. Since the corpus is not part-of-speech tagged, searches cannot be restricted to a particular part of speech. For example, the search for the verb /kòhâykàət/ did not return only the verb, but also the related noun /kaan kòhâykàət/. This is because /kaan kòhâykàət/ is segmented in the corpus as two separate words, despite the fact that it is typically considered a single complex word consisting of /kaan/, a grammatical particle that nominalises a verb, and /kòhâykàət/. In fact, the word segmentation criteria used in the TNC do not seem to be consistent (see section 3.2.3). For example, whereas /khwaamkhlânwăy/ is considered one word, /kaan khlânwăy/ is segmented as two separate words, although both /khwaam/ and /kaan/ are nominalising particles. These disadvantages are likely to be much reduced under Sinclair's approach, where the study of semantic prosody relies on examination of concordance lines.

In sum, we have seen that whereas /kòhâykàət/ and /chòp/ were found to display a negative semantic prosody, /kreençay/ does not show a particular tendency to co-occur with positive or negative words. We have also seen some limitations in this approach to semantic prosody based on looking at statistical collocates.

4.4.2 Sinclair's approach

4.4.2.1 /kreeŋcay/ 'considerate'

To identify the major patterns of /kreeŋcay/, I started by sorting the concordance lines on one position to the left and looking for repeated words in this position. The reason I mainly looked at the left contexts was that there appeared to be more consistent and interesting patterns there than among the right contexts. Using the immediately preceding words as a starting point, I looked for possible patterns. Of the 200 concordance lines, 196 were suitable for analysis: two instances were repetitions⁹, and in two instances, /kreeŋcay/ was being referred to as a word.

The first pattern that emerged has the linking particle /kô/ immediately to the left of the node. /kô kreeŋcay/ occurs 10 times in total. To the left of that expression can be observed one consistent pattern:

/cà/ [verb] ([object/adverb]) /kô kreeŋcay/
/yàak/
/yàak cà/

/cà/ is a challengeability marker; /yàak/ and /yàak cà/ both mean 'want to'. Between the verb and /kô kreeŋcay/, there may be an object or an adverb, depending on the nature of the verb. There are six instances of this pattern altogether. One example from the concordance is illustrated below¹⁰.

⁹ Such repetitions result from double-inclusion of a text in the TNC by error.

¹⁰ The abbreviations used in line 2 of Thai examples are listed in Appendix 3.

Example 4.1

cà	dàa	pen	phaasǎa	ʔaŋkrit	kô	kreeŋcay
CM	curse	COP	language	English	LP	considerate
mèem	ʔàmeerikan	thîi	yùu	klây	klây	
lady	American	SBR	stay	close	close	

‘(I wish) to curse in English, but I feel considerate to this nearby American woman.’

The speaker in the example wishes to curse in English, but out of consideration for the American woman, refrains from doing so. The causal relationship is, however, not stated explicitly. This implied meaning of refraining from performing an action due to consideration for others exists in all examples of this pattern.

Example 4.1 and one other instance also contain a reference to the person who the consideration is for. However, the same pragmatic meaning is also present in four instances where the “person” is not mentioned after /kreeŋcay/, e.g. Example 4.2.

Example 4.2

yàak	khǒɔ	khànǔn	ʔòɔn	maa	tôm	cím	cèew
want	ask	jackfruit	unripe	ASP	boil	dip	sauce
kô	kreeŋcay						
LP	considerate						

‘(I) want to ask for some unripe jackfruit to make some sauce, but I feel considerate.’

In this example, the speaker wishes to ask for jackfruit, but out of consideration (for an unspecified person), he/she does not do so.

In both examples above, there is not any word in the original Thai text that means ‘but’. However, the meaning of ‘but’ can be implied from context, and it is necessary for an accurate translation. There exists one instance, however, where the word /tèɛ/ ‘but’ is present, shown in Example 4.3.

Example 4.3

yàak	cà	chuan	khǎw	pay	thủ	lǎn	dûay
want	CM	invite	3SG	go	scrub	back	with
tèe	kô	kreeŋcay					
but	LP	considerate					

‘(I) want to invite him to go and scrub our backs together, but I feel considerate.’

Therefore, it can be said that /kô kreeŋcay/ is the core of the following extended unit of meaning:

/cà/ [verb] ([object/adverb]) (/tèe/) /**kô kreeŋcay**/ ([person])
 /yàak/
 /yàak cà/

This unit of meaning has the semantic prosody of ‘refraining from performing an action due to consideration for other(s)’. This meaning cannot be derived from any individual word in the sequence; there are no individual words that mean ‘refraining from’ or that show negation. Rather, it is spread across the whole sequence and is a pragmatic interpretation.

The second pattern has the modal auxiliary /tôŋ/ ‘must/have to’ in the first position to the left of the node. Of the 12 instances of /tôŋ kreeŋcay/, 7 are immediately preceded by the negator /mây/. /mây tôŋ kreeŋcay/ means literally ‘do not have to be considerate’. Six of these seven examples function as imperatives, in which case a better translation would be ‘don’t worry, it’s no trouble’. These imperatives always refer implicitly to an imposition expressed in a preceding sentence, as illustrated in Example 4.4.

Example 4.4

thâa	klaaŋkhuun	dùkdùun	păa	kòət	pen	ʔaray
if	night	late	dad	happen	COP	REL
thoo	riak	ʔôo	dây	lăay	ná	
ring	call	Or	POT	PP	PP	
mây	tôŋ	kreeŋcay				
NEG	must	considerate				

‘If at night something happens to you, Dad, feel free to ring me. Don’t worry, it’s no trouble.’

The expression /mây tôŋ kreeŋcay/ is bound up with the previous sentence, in which the speaker encourages the interlocutor to feel free to ring her. It functions to reassure the interlocutor that the action of ringing the speaker would not be an imposition. This function is found in another three instances. One is given as Example 4.5.

Example 4.5

mii	ʔaray	kô	khian	maa	bòok	khàatlăa	ʔaray
have	REL	LP	write	ASP	tell	lack	REL
mây	tôŋ	kreeŋcay	ná				
NEG	must	considerate	PP				

‘If you need anything, just write to me. It’s no trouble at all.’

There are two instances, however, where the sentences preceding /mây tôŋ kreeŋcay/ do not explicitly encourage the interlocutor to perform an action, shown in Examples 4.6 and 4.7.

Example 4.6

rápprathaan	ʔaray	maa	răuyay	khá
eat	REL	ASP	Q	SLP
mây	tôŋ	kreeŋcay	ná	khá
NEG	must	considerate	PP	SLP

‘Have you eaten anything yet? Don’t worry, it’s no trouble.’

In contrast to the previous examples, in this example, the speaker asks the interlocutor a question. However, the question here is an implicit invitation, so the overall pragmatic function is the same.

Example 4.7

súu	hây	súu	hây	
buy	give	buy	give	
mây	tôn	kreeŋcay	ròk	
NEG	must	considerate	PP	

‘(I) will buy it for you. (I) will buy it for you. Don’t worry, it’s no trouble.’

Here, the speaker offers (twice!) to buy a product for the interlocutor. The speaker’s use of /mây tôn kreeŋcay/ encourages the interlocutor to accept the offer, again expressing the pragmatic meaning of not refraining from an imposition.

There are three instances (Examples 4.5, 4.6 and 4.7) where /mây tôn kreeŋcay/ is immediately followed by one of the pragmatic particles /ná/ and /ròk/. /ná/, when used with imperatives, requests or encourages compliance, whereas /ròk/ “is used to counter argue or correct an assumption that an addressee has” (Iwasaki and Ingkaphirom 2005: 190, 195, 201).

From these observations, it might be argued that /mây tôn kreeŋcay/ is the core of the following extended unit of meaning:

[imposition of hearer on speaker] | /mây tôn kreeŋcay/ (ná/)
(ròk/)

In all the examples, the unit of meaning stretches into the previous sentence and encourages the interlocutor to perform an action or accept an offer. This action or offer is normally an imposition on the speaker. By saying /mây tôn kreeŋcay/, the speaker attempts to remove the imposition implied in the previous sentence, reassuring the

interlocutor that it will not be a bother. Thus, this extended unit of meaning has the pragmatic function of ‘reduction of imposition’ – which is thus its semantic prosody.

Let us return to the shorter sequence /t^hoŋ kreeŋcay/ ‘have to be considerate’. Without the negator, it occurs five times altogether. Looking at the extended context, I found that in all instances, the person that the consideration is for is for some reason important or has authority over the person who has to be considerate. One instance is illustrated in Example 4.8, where the speaker, who is an international student, has to be considerate to the host family.

Example 4.8

t ^h oŋ	kin	ʔaahǎan	th ⁱ	khǎw	tham	hây	taam
must	eat	food	SBR	3PL	do	give	follow
weelaa	túk	wan	t ^h oŋ	kreeŋcay	weelaa	cà	khâw ʔòok
time	every	day	must	considerate	time	CM	enter exit

‘(I) have to eat the food they cook for me at the same time every day. (I) have to be considerate when I enter or leave (the house).’

Other than this, /t^hoŋ kreeŋcay/ does not seem to have as clear a pragmatic meaning as /mây t^hoŋ kreeŋcay/. /mây t^hoŋ kreeŋcay/ literally means ‘do not have to be considerate’, but is pragmatically used to reduce an imposition. /t^hoŋ kreeŋcay/, on the other hand, does not seem to have a clear pragmatic function beyond its literal meaning of /t^hoŋ/ ‘must/have to’ plus /kreeŋcay/ ‘considerate’. Moreover, unlike /mây t^hoŋ kreeŋcay/, which requires a preceding sentence that implies an imposition, /t^hoŋ kreeŋcay/ seems to be able to appear anywhere as long as it satisfies all relevant grammatical constraints. Therefore, I would argue that /t^hoŋ kreeŋcay/ does not seem to form part of an extended unit of meaning in Sinclair’s sense. However, it *can* be argued that there is a colligational link between /t^hoŋ/ and /kreeŋcay/; as a verb, we would expect /kreeŋcay/ to colligate with a modal auxiliary, and vice versa.

The final pattern has the negator /mây/ immediately to the left of /kreeŋcay/. /mây kreeŋcay/ ‘without consideration for/inconsiderately’ occurs 27 times in total. The position immediately to the left of the expression is dominated by /yàaŋ/, /bèep/ and /dooy/, which are grammatical particles that create an adverbial clause. There are 17 instances of this pattern altogether. In all these instances, /yàaŋ/, /bèep/, or /dooy/ followed by /mây kreeŋcay/ modifies the verb of the containing clause. Many of the verbs in question refer to unpleasant actions, such as /hũaró sám̄təəm/ ‘laugh insultingly’, /tākoon hækpàak/ ‘yell out’, /thĩaŋ hũa chon fãa/ ‘wrangle’, and /sùup phôn khwan pũy pũy/ ‘puff out smoke’, as shown in Example 4.9 and 4.10.

Example 4.9

khon	lên	kiitâa	tākoon	hækpàak		
person	play	guitar	shout	yell		
sĩaŋ	daŋ	yàaŋ	mây	kreeŋcay	khay	
sound	loud	AZP	NEG	considerate	REL	

‘The man who played the guitar inconsiderately yelled out.’

Example 4.10

khãw	klàw	náamsĩaŋ	yen	phlaaŋ	yìp	bùrii
3SG	say	tone	cold	while	pick	cigarette
nay	kràpãwsãa	maa	sùup	phôn		khwan
in	pocket	shirt	ASP	smoke	puff.out	smoke
pũy	pũy	yàaŋ	mây	kreeŋcay	cãawkhốŋ	hốŋ
puff	puff	AZP	NEG	considerate	owner	room

‘He spoke in a cold voice, reaching for a cigarette in his pocket. He kept on puffing out smoke without consideration for the owner of the room.’

There are a few neutral or even positive verbs in this context, such as /lóm tua loŋ nồŋ/ ‘lie down’ and /yím khwãaŋ thũŋ bayhũu/ ‘grin from ear to ear’. However, in context these refer to actions unpleasant to another party, as can be seen from Example 4.11.

Example 4.11

pêe yím khwâaŋ thǎŋ bayhǔu yàaŋ mây
 Pay grin broad till ear AZP NEG
 kreeŋcay khon nâa pen tùut thîi nâŋ yùu
 considerate person face COP ass SBR sit ASP
 khâaŋ khâaŋ
 next.to next.to
 ‘Pay grinned from ear to ear, without consideration for the person sitting
 next to him who was frowning.’

Within this context, it might be said that /mây kreeŋcay/ is the core of the following extended unit of meaning:

[action inconsiderate to another] /yàaŋ/ /mây kreeŋcay/ ([person])
 /bêeŋ/
 /dooy/

This unit of meaning could thus be argued to have a semantic prosody expressing ‘disapproval of behaviour’. That is, use of the unit is motivated by the speaker’s wish to express disapproval of an action encoded by the verb of the containing clause which they deem inconsiderate.

We have seen from the above analyses that /kreeŋcay/ is part of the core of (at least) three different extended units of meaning. Across the 196 instances, three patterns emerged, as summarised in Table 4.2.

Extended units of meaning		Frequency of occurrence
/cà/ [verb] ([object/adverb]) (/têe/) /kâ kreeŋcay/ ([person]) /yàak/ /yàak cà/		6
[imposition of hearer on speaker] /mây tǔŋ kreeŋcay/ (/ná/) (/ròk/)		6
[action inconsiderate to another] /yàaŋ/ /mây kreeŋcay/ ([person]) /bêeŋ/ /dooy/		17

Table 4.2 Frequency of occurrence of each extended unit of meaning containing /kreeŋcay/

In order to make sure that I did not leave out any extended units of meaning, I examined the remaining 167 concordance lines again, scrutinising both the left and the right contexts this time. In all 167 instances, I found that /kreeŋcay/ is used as a straightforward verb meaning ‘(to) be considerate’, and does not form part of any discernible fixed pattern. However, as a transitive verb of cognition, it does have the colligations and semantic preferences one would expect of a verb of this type. For example, it colligates with a subject and an object, which in turn have a semantic preference to be human. However, there are many instances where the subject is implicit, and a number where the object is not present, although context normally implies straightforwardly whom the consideration is for. /kreeŋcay/ also has the colligations that are generally characteristic of any kind of verb. For instance, it colligates with modal auxiliaries, as in the previously discussed case of /tôŋ kreeŋcay/. It also colligates with /khwaam/ to create a noun, /khîi/ to create an adjective, and /yàaŋ/ to create an adverbial clause, these being grammatical particles which with any Thai verb may co-occur.

4.4.2.2 /kòŋhâykàət/ ‘cause’

To identify the patterns of /kòŋhâykàət/, I mainly examined the right contexts of the node, as there appeared to be much more consistent and interesting patterns on this side. There was one repeated concordance line, leaving 199 instances for analysis.

Of the remaining 199 instances, there are 9 instances where /kòŋhâykàət/ is nominalised; that is, it is preceded by /kaan/, the grammatical particle that creates a noun. Of the remaining 190 instances, my observation of the right contexts reveals as many as 186 instances where /kòŋhâykàət/ is used as the verb of a clause and is immediately

followed by a noun. In two other instances, /kòhâykəət/ is followed by a series of verbs and a noun. In one of the remaining two instances, the word that follows /kòhâykəət/ is a verb. However, as a native speaker, I could not read it as a coherent text. The verbal reading in this case seems to have arisen from data corruption in the corpus. Therefore, I ignored this instance. In the final instance, /kòhâykəət/ is followed by the aspect marker /khân/. The fact that /kòhâykəət/ is mostly followed by a noun does not come as a surprise, considering the fact that /kòhâykəət/ is a transitive verb and thus would be expected to have an object after it. Interestingly, 174 of these nouns refer to abstract concepts. 74 of these abstract nouns are formed by the nominalising particles /kaan/ and /khwaam/, such as /kaan plianplæŋ/ ‘change’, /kaan phátthánaa/ ‘development’, /khwaam lamʔiaŋ/ ‘bias’, and /khwaamkhàtyéŋ/ ‘conflict’.

In terms of semantics, of the 188 noun tokens that follow /kòhâykəət/, 107 can be considered negative. The three most frequently co-occurring negative nouns are /panhăa/ ‘problem’, /khwaamsǎhăay/ ‘damage’, and /ʔantàraay/ ‘danger’, with 14, 12, and 8 examples respectively. 46 other noun tokens are positive. /pràyòt/ ‘benefit’ is the most frequent positive noun (8 times). There are 35 instances where the following nouns are neutral in meaning.

Some of these object nouns can be categorised into more specific semantic categories. The words with a negative meaning are in bold, and those with a positive meaning are underlined. Brackets show numbers of occurrences greater than one.

Health:

/rôok/ ‘disease’ (3), **/máreŋ/** ‘cancer’, **/ʔaakaan wiconcàrit/** ‘insanity’, **/ʔaakaan sàŋ/** ‘tremor’, **/ʔaakaan ʔàksèp/** ‘inflammation’, **/ʔaakaan phéɛ/** ‘allergy’, **/sǎw/** ‘acne’, **/phaawá ruu khǔm khǒn ʔùt tan/** ‘blocked pore’, **/kaan ràkhaaykhəaŋ/** ‘irritation’, **/khwaam cèppuət/** ‘pain’

Difficulty:

/panhãa/ ‘problem’ (14), /ʔupàsàk/ ‘obstacle’, /wikrit/ ‘crisis’, /khwaam yũnyâak/ ‘difficulty’, /khwaam sàpsõn/ ‘confusion’, /khwaam wũnwaay/ ‘disorder’

Danger:

/ʔantàraay/ ‘danger’ (8), /phay/ ‘hazard’ (3), /pháyantàraay/ ‘danger’

Absence:

/khwaam máy pen tham/ ‘unfairness’, /khwaam ráy rábiap/ ‘disorder’, /khwaan ráy sàthianráphâap/ ‘instability’

Benefit:

/pràyòot/ ‘benefit’ (8), /phõnpràyòot/ ‘benefit’ (2), /phõntòoptɛɛn/ ‘reward’

Change of state:

/kaan plianplɛɛŋ/ ‘change’, /kaan khlũanwǎy/ ‘movement’ (3), /khwaam plianplɛɛŋ/ ‘change’, /kaan tham hây sǎa pay/ ‘losing’, /khwaam sùam/ ‘decline’, /khwaam sùamsoom/ ‘deterioration’, /kaan phátthánaa/ ‘development’

Damage:

/khwaamsǎahǎay/ ‘damage’ (12), /phõn sǎahǎay/ ‘damage’ (2), /phõnkràthóp/ ‘effect’ (6), /phõnsǎa/ ‘negative effect’ (2) /phõn khâaŋkhiaŋ/ ‘side effect’

Conflict:

/khwaam tɛɛkyɛɛk/ ‘disharmony’, /khwaamkhàtyéɛŋ/ ‘conflict’, /kaan kràthóp kràthân/ ‘conflict’

Cognition/emotion:

/khwaam khâwɔay/ ‘understanding’, /khwaam bandaan cay/ ‘inspiration’, /khwaam miì caykwâaŋ/ ‘generosity’, /khwaam phũphɔɔcay/ ‘satisfaction’, /rɛɛŋ bandaan cay/ ‘inspiration’, /rɛɛŋ cuŋcay/ ‘motivation’, /khwaam kreenɔay/ ‘consideration’.

Examples 4.12 and 4.13 show the use of /kòohâykòət/ in context.

Example 4.12

mũa	khun	kròot	khwaamrúusùk	thi	mák	cà
when	3SG	angry	feeling	SBR	often	CM
kòohâykòət	panhãa	kô	khæ	khwaam	kròot	
cause	problem	LP	COP	NMLZ	angry	

‘When you are angry, the feeling that often causes you a problem is anger.’

Example 4.13

khûn	chôp	râaŋ	phàconphay	tham	ʔàray	thîi
3SG	like	story	adventure	do	REL	SBR
kòohâykàət		ʔantàraay				
cause		danger				

‘You were adventurous and liked to do things that caused danger.’

Examining the left contexts, I found that many of the subjects of /kòohâykàət/ are abstract nouns. Many of these nouns are, again, formed by the grammatical particles /kaan/ or /khwaam/, such as /kaan namkhâw/ ‘import’ and /khwaamrúusùk/ ‘feeling’. There are also cases where the subject is a pronoun making general reference to the preceding clause(s). However, this pronoun is omitted, leaving only the preceding verb clause(s). These two types of subjects are linked, because they both involve a subject whose reference is the general situation under discussion. There are also a few concrete noun subjects, such as /sǎaŋ/ ‘sound’ and /yaa/ ‘medication’.

We have thus seen that /kòohâykàət/ has a semantic preference for abstractness. It also has a colligation for nouns, especially those beginning with two particular grammatical particles (nominalizers). These requirements apply to both the subject and the object of /kòohâykàət/. Since the grammatical particles which /kòohâykàət/ attracts are abstract noun-forming particles, we can say that the semantic preference and the colligation are linked here. Despite these associations, I would argue that the pattern in which /kòohâykàət/ is regularly used, that is, an abstract noun subject followed by /kòohâykàət/ followed by an abstract noun object, is not an extended unit of meaning in Sinclair’s sense. Rather, /kòohâykàət/ is used as a unit of meaning on its own in these examples. Its requirements for a subject and an object noun colligation are just the general requirements one would expect of any transitive verb, although in this case the

object nouns tend to be (negative) abstract concepts. More importantly, this pattern does not have a clear pragmatic function that is distinct from its literal meaning. Thus, the single-word unit of meaning has colligations and semantic preferences, but not a semantic prosody beyond its base meaning.

4.4.2.3 /chôp/ ‘like’

To identify the patterns around /chôp/, I examined the right contexts of the node, as the patterns on this side seemed to be more consistent. Of the 200 instances, 184 were suitable for analysis. Of the rest, /chôp/ is a person’s name in two; in 11, /chôp/ means ‘righteous’ (a homophone); in one, /chôp/ is part of a proverb; and in two instances, /chôp/ is part of a compound noun.

More than half of the words immediately to the right of the node are nouns or verbs. There are 55 noun tokens and 66 verb tokens. Table 4.3 illustrates distribution of the co-occurring noun and verb tokens across meaning categories.

	Positive	Negative	Neutral	Total
Pattern with noun complement	9	8	38	55
Pattern with verb complement	2	27	37	66

Table 4.3 Distribution of co-occurring noun and verb tokens across meaning categories

From this distribution, it might be said that when /chôp/ is followed by a noun, it does not have a tendency to co-occur with a positive or negative word in particular. There is little difference in proportion between the positive and negative nouns, although the majority are neutral. However, this does not seem to be the case when /chôp/ is followed by a verb (to create a serial verb structure). There are 27 instances where /chôp/ is followed by a negative verb, but only 2 where the verb is positive. The negative verbs

include /biatbian/ ‘take advantage of’, /klêɛŋ/ ‘tease’, /khùu/ ‘threaten’, /wícaan/ ‘criticise’, and /pân rûaŋ/ ‘make up stories’. Examples 4.14 and 4.15 exemplify these verbs in context.

Example 4.14

thoŋ	sàay	nâa	bòk	wâa	phûuyŋ	phan	ní
Tong	shake	face	tell	COMP	woman	kind	DEM
chôp	pân	rûaŋ	hây	tuaʔeeŋ	lúkláp	nâa	kónhăa
like	make	story	CAUS	REFL	mysterious	PFX	search

‘Tong shook her head, saying that this kind of women likes to make up stories to make themselves intriguing.’

Example 4.15

khăw	kliat	phûu	pen	phô	maa	naan
3SG	hate	person	COP	father	ASP	long
phró	chôp	thamráay	mêe			
because	like	hurt	mother			

‘He/she has hated his/her father for a long time, because he likes to hurt his/her mother.’

Looking at the left contexts of the 121 instances where /chôp/ is followed by a verb or a noun complement, I found that the majority of the subjects of /chôp/ are human beings. There are two instances where the subject is a collective noun, such as ‘nation’ or ‘government’, and one where the subject is some dogs.

Thus, we see that /chôp/ frequently appears in the pattern of a subject plus /chôp/ plus a noun or a verb complement. However, I would argue that when /chôp/ is followed by a noun complement, it is used as a single-word unit of meaning – that is, as a straightforward verb meaning ‘like’. The overall sequence of /chôp/ plus a noun complement does not have a clear pragmatic meaning beyond the literal meaning of ‘someone liking a thing’. The /chôp/ unit of meaning is, however, a transitive verb of cognition, and therefore has requirements for its (nominal) complement and for its

subject, which has to be a human being or at least a conscious being, these being the general colligation and semantic preference characteristics of a verb of this type.

However, arguably /chôp/ forms part of an extended unit of meaning in Sinclair's sense when it is followed by a verb, as follows:

[person] /**chôp**/ [verb] ([object/adverb])

Here, /chôp/ is the core of the extended unit of meaning. This unit colligates with a verb complement. Unlike the pattern where /chôp/ is followed by a noun complement, this extended unit has a pragmatic meaning beyond its literal meaning of 'someone liking something'. In all the examples, the reference is to a *habit* of the clause subject. This unit of meaning is frequently associated with *bad* habits or neutrally-evaluated habits, and only rarely associated with *good* habits. This reference to a habit is very rarely, if ever, present in the pattern where /chôp/ is followed by a noun complement. Bad habits attributed to the clause subject are expressed in both Examples 4.14 and 4.15 above: the habit of hurting and the habit of making up stories respectively. This is thus the unit's semantic prosody, since the meaning of habit is not implied individually by either /chôp/ or the verb that follows it. We can see, for instance, that "likes to hurt (someone)" does not directly imply that this is a (bad) habit, in the sense of a recurring behaviour, but in Thai, this meaning is present. This pragmatic meaning is spread across the unit.

There are two instances where /chôp/ is followed by a verb complement, but not immediately. Instead, the verb complement is preceded by the complementizer /thîi/ and the challengeability marker /cà/, as in Example 4.16 (which also exemplifies the rarer good habits).

Example 4.16

sàmátchaa phromsiri pen nùm thîi mii rɔɔyým ʔòpʔùn
 Samatcha Promsiri COP man SBR have smile warm
 chòp thîi cà duulɛ sùkkhàphâap fan khɔɔŋ
 like COMP CM look.after health tooth of
 tuaʔeeŋ pen yàaŋ dii
 REFL COP AZP good
 ‘Samatcha Promsiri is a man who has a warm smile and likes to take good
 care of his teeth.’

Here, /chòp/ arguably also forms part of an extended unit of meaning. The unit also has the meaning of a habit. We can therefore revise the previously presented extended unit of meaning whose core is /chòp/ as follows:

[person] /**chòp**/ (/thîi cà/) [verb] ([object/adverb])

Of the other 61 instances, there are four instances where /chòp/ is followed by a clause preceded by the causative marker /hây/, which literally means ‘give’. This sequence forms a causative serialisation that consists of a causing and a resulting situation (Iwasaki and Ingkaphirom 2005: 239), as shown in Example 4.17.

Example 4.17

ran tɔɔŋ phêəm náamnàk léew ná
 Run must increase weight ASP PP
 ʔiŋ chòp hây ran kêem yúy yúy
 Ing like CAUS Run cheek chubbychubby
 ‘Run, you must increase your weight. I want you to have chubby cheeks.’

It is arguable that /chòp hây/ is the core of an extended unit of meaning with the pragmatic function of expressing a desire for someone to do something. In Example 4.17, the speaker’s desire for Run to have chubby cheeks is expressed across the whole unit rather than by any individual word of the unit. This extended unit of meaning can be laid out as follows:

[person] /**chòp hây**/ [another person] [verb] [(object/adverb)]

Of the remaining 57 instances, there are five where /chôp/ appears in a fixed sequence; it is preceded by /taam/ ‘follow’ in /taam chôp/. There are also two instances where /cay/ ‘heart’ or /têe/ ‘but’ occurs between /taam/ and /chôp/ in /taam cay chôp/ and /taam têe chôp/ respectively. These three fixed sequences all literally mean ‘as you please’. Pragmatically, they are politeness markers expressing deference to someone else’s choices. Thus, they might be argued to constitute a set of variants of one single extended unit of meaning as follows:

/taam/ (/cay/ /chôp/
 (/têe/)

Example 4.18

kày	yâaŋ	thîi	nîi	pen	kày	yâaŋ	thîi
chicken	grill	at	DEM	COP	chicken	grill	SBR
hêeŋ	nûm	mây	chum	náamman	səəp	próom	cèew
dry	tender	NEG	soak	oil	serve	together	sauce
lê	náamcîm	hây	lûak	taam	chôp		
and	dip	CAUS	chose	follow	like		

‘The grilled chicken here is dry and tender. It is not oily. It is served with a variety of dips for you to choose as you please.’

In the remaining 52 instances, /chôp/ is also used as a verb meaning ‘like’, but no frequent patterns stand out clearly. Here, /chôp/ is followed by a variety of types of object. In 41 instances, there is no explicit object after /chôp/. Rather, the thing liked can be inferred from context (mostly the preceding discourse). In nine instances, /chôp/ is followed by a pronoun (all but one referring to human beings). In two instances, /chôp/ precedes an object clause beginning with /thîi/ (which is a complementizer) followed by a subject and a verb.

4.4.2.4 Discussion

The concordance analysis of /kreeŋcay/ and /chôp/ has allowed me to identify some discernible extended units of meaning in Sinclair's sense. These units of meaning, as we have seen, each have specific colligations and semantic preferences as well as a clear pragmatic function that is distinct from the literal meaning of the core of the unit.

Nevertheless, there are also many instances where /kreeŋcay/ and /chôp/ seem to be used as a unit of meaning on their own. In fact, this is the major (most frequent) use of both /kreeŋcay/ and /chôp/. /kòhâykàt/ seems to be used as a unit of meaning on its own in all instances. In these examples, the verbs' overall pragmatic function is not easily distinguishable from their literal semantics. Moreover, they appear in a great variety of contexts limited only by grammatical restrictions. In these examples, arguably, the unit of meaning is not extended, counter to Sinclair's view that most units extend across more than one word (see section 2.3.2.1). Rather, we do actually see a single word that is a carrier of meaning largely independent of the words around it.

However, when /kreeŋcay/, /kòhâykàt/ and /chôp/ are used independently, they *do* still require the general colligations and semantic preferences that are characteristic of the class of words they belong to. As transitive verbs of cognition, /kreeŋcay/ and /chôp/ have an object and (at least) a conscious-being subject. /kòhâykàt/, as a transitive verb, requires a subject and an object, both of which have a semantic preference that tends towards abstractness. These associations between /kreeŋcay/, /kòhâykàt/, and /chôp/ and their colligations and semantic preferences may not create an extended unit of meaning in Sinclair's sense, but they do constitute a specific pattern in which /kreeŋcay/, /kòhâykàt/, and /chôp/ are used. The patterns of a word, as Hunston and Francis (2000:

37) define the term, are made up of “all the words and structures which are regularly associated with the word and which contribute to its meaning.”

My analysis of the semantic prosodies presented above in section 4.4.2 may invite the same criticism as Tognini-Bonelli’s analysis of *in the case of* discussed in section 2.7.7. Particularly, from the polarity-oriented perspective, it might be argued that the semantic prosody of “deference to someone else’s choice” for the unit /taam cay chôp/ and the semantic prosody of “someone expressing a desire for someone else to do something” for the unit whose core is /chôp hây/ are unusual, or perhaps invalid, as they do not clearly incorporate expressions of positive or negative evaluation. That said, I maintain that the semantic prosodies in question are not unusual or wrong from the perspective of the EUM-oriented approach. This issue will be later explored in detail in Chapter 5 in section 5.6.3.

4.5 General discussion

As my discussion above illustrates, both the polarity-oriented approach and the EUM-oriented approach proved viable routes to investigate the semantic prosody of the words under investigation. Using the polarity-oriented approach, I was able to identify some semantic prosodies. While /kòhâykàət/ and /chôp/ were found to display a negative semantic prosody, /kreeŋcay/ was found to have a neutral semantic prosody, as it does not tend to co-occur particularly with positive or negative words. Using the EUM-oriented approach, on the other hand, I was able to identify some extended units of meaning around /kreeŋcay/ and /chôp/ and these units’ pragmatic function. The EUM-oriented approach does not discover any extended units of meaning around /kòhâykàət/,

however. What we have found is that /kòòhâykàèt/ is *always* used independently as a single-word unit of meaning, at least in the 200 random examples that I looked at.

Thus, the preceding analysis has allowed me to demonstrate the differences between Sinclair's approach and Louw, Stubbs, and Partington's approach in great depth. In terms of methodology, the Sinclairian approach relies on concordance analysis. To identify the semantic prosody of the extended unit of meaning around /kreeŋcay/, for instance, I had to examine its extended co-text. This enabled me to identify extended units, such as /cà/ [verb] ([object/adverb]) (/tèè/) /kô kreeŋcay/ ([person]), with its semantic prosody of 'refraining from performing an action due to consideration for other(s)'. This EUM-oriented method discovers not only semantic prosody, but also colligation and semantic preference. Thus, under this Sinclairian approach, semantic prosody cannot be discussed independently of colligation and semantic preference, as Stubbs notes in a discussion of "semantic schemas" (i.e. units of meaning):

These semantic schemas can be modelled as clusters of lexis (nodes and collocates), grammar (colligation), semantics (preferences for words from particular lexical fields), and pragmatics (connotations or discourse prosodies).

(Stubbs 2001: 96)

By contrast, Louw, Stubbs and Partington's approach generally relies on collocate analysis. These scholars consider semantic prosody as a word's tendency to co-occur with positive or negative words. To identify the prosodies of /kreeŋcay/, /kòòhâykàèt/, and /chôòp/, all I needed to do was to examine whether they tend to co-occur with positive or negative collocates. /kòòhâykàèt/ and /chôòp/ were found to display a negative semantic prosody. /kreeŋcay/ does not display a positive or negative semantic prosody, as there is little difference in the proportion of positive and negative collocates.

The analysis has allowed me to demonstrate in a very clear and concrete way the differences in both methodology and underlying concept between the two approaches. In fact, these differences have recently been made obvious by Partington's proposal to change the terminology to evaluative prosody. Evaluative prosody, Partington (2014: 283) argues, is a word's *inherent* evaluative potential to co-occur with other items of the same evaluative polarity. Exactly this kind of evaluative potential is evident for /kòchâykàət/ and /chôp/ on the basis of the collocate analysis. Having a negative evaluative prosody, /kòchâykàət/ and /chôp/ tend to occur in negative environments to maintain evaluative harmony in the discourse. For example, the collocate analysis shows that /chôp/ habitually co-occurs with negative verbs, such as /sòtrúusòthěn/ 'snoop' and /waanthâa/ 'act big', whereas /kòchâykàət/ frequently co-occurs with negative nouns, such as /phönsǎa/ 'negative effect' and /khwaamsǎahǎay/ 'damage'.

In sum, then, my answer to the research question of "what are the advantages and disadvantages of the major approaches to semantic prosody proposed in the literature for describing semantic prosody in Thai?" would be as follows:

Both approaches are useful for different purposes. The polarity-oriented approach is useful when one's aim is to examine a word's tendency to appear in an evaluatively positive or negative context. Particularly, it reveals the implicit evaluation of a word whose evaluative potential is not immediately obvious from its core semantics, as we have seen with /kòchâykàət/ and /chôp/. The knowledge obtained from this type of analysis will thus be useful for scholars interested in study of evaluation in discourse. It will also be beneficial for those who wish to exploit a semantic prosody for stylistic effects (see sections 2.3.1.1 and 2.3.1.2). That said, this type of analysis of semantic

prosody also has a disadvantage. That is, semantic prosody identified through this approach is limited to the positive vs. negative opposition rather than express a variety of expression of evaluation. Unlike the polarity-oriented approach which reveals only the implicit evaluation of a word, the EUM-oriented approach gives us more details about the patterns in which the word occurs. We have seen that this approach discovers not only semantic prosody, but also colligation and semantic preference. Moreover, within this approach, semantic prosody can be any pragmatic function or meaning, and is not confined to the positive vs. negative opposition or expression of evaluation or attitude. However, this approach also has a limitation. As its name suggests, the EUM-oriented approach is only applicable when we work under Sinclair's theory of language. Scholars who adopt the polarity-oriented approach or who do not work within Sinclair's framework of model of the extended unit of meaning may find it of little use (see section 2.7.7).

They both also have some advantages and disadvantages in terms of practicality. We have seen in the case of /kəʊhâykət/ 'cause' that the analysis of collocates alone is sometimes not sufficient to access semantic prosody (see section 4.4.1.4). So, this is a disadvantage of the polarity-oriented approach. Regarding the EUM-oriented approach, it should be obvious from this study that it is much more time-consuming than the polarity-oriented approach. It also requires much more space to present the results. So, this approach may not be practical if it is necessary to investigate a large number of words. Furthermore, it uses less actual data than the polarity-oriented approach. We have seen that with the EUM-oriented approach, due to time constraints, only 200 concordance lines per word could be analysed. With the polarity-oriented approach, on the contrary,

all the corpus data was used; the statistical collocates of the word under investigation were automatically generated from the whole corpus. Despite these drawbacks, the EUM-oriented approach produces outcomes that cannot be accessed via the polarity-oriented approach, as previously discussed. These practical issues as well as the effectiveness of each approach will also be taken into consideration when selecting the optimal approach for the remaining two research questions.

It might be objected that these answers to my research question are already obvious in the literature, and that I did not actually have to conduct the analysis to arrive at this point. I would argue that employing both of the approaches to analyse the Thai data has in fact yielded outcomes that I could not have obtained without carrying out the analysis, as follows.

First, the analysis reveals that both approaches do operate in Thai, as they do in English. Even though Thai is not very different from English in terms of syntax, it would theoretically be quite possible for syntactic differences between the languages to have the effect that only one of the approaches operates, whereas the other does not. That both approaches do operate in Thai proves the cross-linguistic validity of both.

Second, the analysis reveals that the approaches produce the same kind of results both in English and in Thai. Employing the Sinclairian approach, I identified some discernible extended units of meaning in Thai, as we have seen with /kreenɕay/ and /chôɔp/, just as Sinclair and Hunston find with, for example, *naked eye* and *budge* (see section 2.3.2.1 and section 2.3.2.2). In addition, in the cases where the words are used independently as a unit of meaning on their own, the concordance analysis leads to a characterisation of the very general patterns in which they are used. This is similar to

what Hunston and Francis (2000) arrive at in terms of Pattern Grammar. (The role of Pattern Grammar in this kind of analysis will be discussed in detail later in section 5.6.1.)

We have so far discussed the advantages and disadvantages of the two major approaches. Particularly, I have argued that these two approaches are useful for different purposes, and that I will consider their effectiveness when selecting the optimal approach for the remaining two research questions. Whereas the polarity-oriented approach reveals the hidden evaluative potential of a word, the EUM-oriented approach gives details about its phraseological behaviour. Therefore, if time is not an issue, and we do not have a specific purpose in mind, applying both approaches to the study of a word will allow us to have a more comprehensive understanding of the word in question. Baker and Egbert (2016: 3) refer to the use of multiple approaches similarly to the combined method that I just suggested above as *methodological triangulation*.

For instance, it is straightforward to argue the results obtained from the polarity-oriented approach, which uses all the corpus data, can help enhance the results gained from the EUM-oriented approach, which uses much less data. For instance, in the analyses of /kòçhâykèət/ and /chôp/, the EUM-oriented approach and the polarity-oriented approach produced much the same results. The EUM-oriented approach reveals that /kòçhâykèət/ is frequently followed by a negative abstract noun. The polarity-oriented approach likewise discovers that the verb tends to occur with negative words, many of which are negative abstract nouns. In the case of /chôp/, the EUM-oriented approach reveals that the verb is frequently followed by a negative verb in a serial verb structure. The sequence of /chôp/ followed by a negative verb forms an extended unit of meaning with a pragmatic function of expressing a bad personal habit. Similarly, the

polarity-oriented approach demonstrates that /chôp/ tends to co-occur with negative verbs. In these cases, then, it is easily arguable that applying both approaches is beneficial, as the results gained from one approach helps to increase the credibility of the results obtained from the other.

That said, we have also seen a case, namely /kreŋcaj/, where the results from the two approaches do not appear to share much in common, if anything. In this case, applying both approaches does not seem to increase the credibility of the results. Rather, as earlier argued, it maximises our understanding of the word's hidden evaluative potential and its phraseological behaviour – but separately.

So we have seen the benefits that might be gained from methodological triangulation. Further studies of semantic prosody in Thai might quite legitimately opt to triangulate in this way. In fact, it might be argued that I have used methodological triangulation in the present analysis, as I have applied the two approaches to study exactly the same set of data. However, I would argue that the use of the two approaches here should not be seen as methodological triangulation, as my aim is to make a *comparison* between the two approaches. If I was, on the other hand, interested primarily in the semantic prosodies of the specific words under study, then my application of the two approaches would be seen validly as methodological triangulation. Moreover, in the remaining two analyses where I have a specific purpose in mind, I will employ one approach over the other rather than triangulate the two approaches. However, there will be one part in Chapter 5 where I will briefly apply the polarity-oriented approach to an analysis which will be mainly conducted from the EUM-oriented perspective (see section 5.6.3); this does constitute triangulation in the strict sense.

4.6 Concluding remarks

In this chapter, I have examined the advantages and disadvantages of the polarity-oriented approach and the EUM-oriented approach to the study of semantic prosody in Thai. To do the analysis, I examined three Thai verbs, using each of the two approaches. The results of the analysis showed that both approaches could be applied without major difficulty to the study of semantic prosody in the Thai language, but that they are useful for different purposes. While the polarity-oriented approach reveals the hidden evaluative potential of a word, the EUM-oriented approach gives us details about the phraseologies where the word occurs. Both approaches also have some advantages and disadvantages in terms of practicality issues (see section 4.5). Having established the benefits of the two approaches and seen their weaknesses, in the next two analysis chapters, I will go on to address my remaining two research questions, which focus on the study of semantic prosody itself, rather than on the methodology.

Chapter 5 – Variation in semantic prosodies across genres

5.1 Introduction

In this chapter, I address the second research question: what variation in semantic prosodies across genres can be identified for Thai words? I will first present my approach to the analysis in section 5.2. Then I will give a brief overview of the genres in the Thai National Corpus in section 5.3. This involves a brief discussion of the term *genre* and also the selection of genres to be included in my analysis. I then go on to introduce the words whose semantic prosody will be explored in section 5.4. In section 5.5, I present the results of the analysis; each word will be separately discussed in detail. A discussion of the results of the analysis is provided in section 5.6.

5.2 Approach to the analysis

The aim of this chapter is to explore what variation in semantic prosodies across genres can be identified for Thai words. In order to carry out the analysis, the EUM-oriented approach was adopted (see section 3.4.2). This approach was selected because my aim is to investigate whether the same word will appear in different or similar (extended) units of meaning, and will have different or similar pragmatic functions, across genres. In the literature, the EUM-oriented approach has previously been used successfully in the analysis of semantic prosody across genres. Hunston (2007: 252), for example, examines the concordance lines of the lemma *cause* and finds that the lemma's usual negative semantic prosody is absent in scientific contexts, where the entities that are caused are neither desirable nor undesirable (see section 2.3.4.2).

In this analysis, as per the method established in section 3.4.2 as well as illustrated in Chapter 4, I examined 200 randomly-selected concordance lines for each selected word in each genre. This means that I examined 800 concordance lines for each word. I identified the major patterns around these words in each genre according to Sinclair's model of the extended unit of meaning, considering colligation, collocation, semantic preference and semantic prosody. In my discussion below, I present each proposed extended unit of meaning in a one-line format using the same notations I used in Chapter 4 (see section 4.2.2 for the notations).

5.3 Genres in the TNC

Genre is a confusing term, which can mean different things to different linguists (Lee 2001: 37). Biber (1988: 170) defines *genre* by distinguishing it from *text type*. For Biber (1988: 170; 1993: 245), *genre* refers to text classification on the basis of external criteria, such as “situations, purposes, and functions of text in a speech community”. *Text type*, on the other hand, refers to text classifications on the basis of internal criteria, such as linguistic forms (Biber 1988: 170). In Biber's sense, it then follows that two texts that belong to different genres may belong to the same text type, if they have similar linguistic characteristics (Lee 2001: 39).

Genre is closely related to *register*. Biber (1993: 244) uses these two terms interchangeably in some of his work, although in some places he distinguishes *genre* from *register* (e.g. Biber and Conrad 2009: 2). Other linguists, such as Martin (1993 cited in Lee 2001: 42), treat *genre* and *register* as two different concepts. However, as the

difference between genre and register is not a central issue in my thesis, I will opt to treat genre and register as the same concept, and use the term *genre*.

The texts in the TNC are categorised into different genres on the basis of external criteria, such as “the purpose of communication, participants, and the settings of communication” (Aroonmanakun 2007: 8). This practice is adapted from Lee’s proposal for categorising texts in the BNC into different genres (Aroonmanakun 2007: 8). Lee (2001: 46) defines a genre as “a grouping [of texts] according to purposive goals, culturally defined.” He contends that texts in a corpus should be classified into genres, as “language teachers and researchers need to know exactly what kind of language they are examining or describing” (Lee 2001: 37).

Each genre in the TNC varies in terms of size. While some genres consist of millions of words, others contain just a few tens or hundreds of thousands of words. To make sure that there will be sufficient data for my analysis, I will examine only those genres that have more than five million words. These genres are: academic writing (9,972,618); fiction (7,960,240); non-academic non-fiction (5,595,242); and newspaper reports (5,452,098). The genres labelled ‘academic’ and ‘non-academic’ both relate to non-fiction texts (Lee 2001: 59). However, the ‘academic’ genre consists of texts that are aimed at audiences at university level, while the ‘non-academic’ genre consists of texts that are aimed at general audiences (Lee 2001: 59). The texts in the ‘non-academic’ genre mainly come from books. The subject matters covered within this genre parallel those in the ‘academic’ genre and include ‘humanities and arts’, ‘medicine’, ‘natural science’, ‘politics, law, and education’, ‘social science’, and ‘technology and engineering’ (Following the BNC’s usage described by Lee 2001: 58-59, as confirmed by

Aroonmanakun in personal communication). Besides these four genres, there are only two other genres that consist of more than one million words: biography (1,777,982) and law (1,360,076). However, I excluded these two genres, as I do not think they are large enough to yield sufficient data for the analysis, and moreover because the large jump in size between the newspaper reports and biography genres from more than 5,000,000 to less than 2,000,000 words furnishes a natural cut-off point. In this chapter, I will use the shorthand labels *academic*, *fiction*, *newspaper*, and *non-academic* when I refer to the four genres under investigation.

5.4 Selection of Thai words

The words I investigate in this chapter will be different from the words I examined by the EUM-oriented approach in Chapter 4. In Chapter 4, I selected a very small number of words which I subjectively deemed interesting; by contrast, in this chapter, I set criteria for word selection. Had I used my own preferences, I might well have been biased towards choosing words that I expected to yield interesting results. The use of formal, quantitative criteria for word selection will also allow me to explore a wide range of words whose translation-equivalents in English or other languages have probably never been studied in the literature before.

The criteria for selecting the words are as follows:

1. Only word types with overall corpus frequency between 10,000 and 12,000 tokens were considered for selection. This range of frequency was selected because it contains a reasonable number of words (78 words). Not many word types have a frequency above 12,000. For example, there are only 87 types

with a frequency between 20,000 and 30,000. In fact, the higher the frequency range, the fewer words appear in that range.

2. Only verbs were considered for selection. Of the 78 word types in the frequency range mentioned above, 27 are verbs. I had two reasons for choosing verbs. First, in Chapter 4, only three verbs were investigated, so it would be worthwhile to analyse more verbs. Second, unlike, say, adjectives which are in many cases inherently evaluative in meaning, verbs are more likely to be neutral in their basic semantics. It will be interesting to see if semantically neutral words form different extended units of meaning with varying pragmatic functions across genres.
3. Only verbs that appear at least 500 times in each different genre were analysed. Of the 27 verbs, 21 meet this last criterion. Two of the 21 needed to be excluded on other grounds. They are /wát/ and /râaŋ/. These verbs are homonyms. Each can be both a verb and a noun and has more than one meaning. /wát/, as a verb, means ‘measure’, and as a noun means ‘temple’. /râaŋ/, as a verb, means ‘draft (a document)’, and as a noun means ‘person’s figure’. Initial examination of the concordances showed that there are only a few instances of /wát/ and /râaŋ/ functioning as a verb across all the genres. /wát/ and /râaŋ/ are therefore excluded from the analysis, as there is not sufficient data to look at. There are thus 19 verbs remaining for the analysis. Table 5.1 shows the 19 verbs, their meanings, and their frequency of occurrence in the whole TNC.

	Verbs	Meanings	Frequency of occurrence
1.	/pràkàat/	announce, declare	11,764
2.	/càp/	touch, arrest, etc.	11,677
3.	/klaaypen/	become	11,626
4.	/hăn/	turn	11,587
5.	/ʔaasăy/	live, rely on, refer to	11,261
6.	/yɔɔm/	resolve, agree	11,246
7.	/yók/	lift, refer, move, give, dismiss, draft (a document)	11,170
8.	/yũn/	stand	11,123
9.	/khûapkhum/	control, supervise, confine	11,121
10.	/níyom/	like	11,009
11.	/waan/	put, be placed, set, walk	10,992
12.	/rɔɔ/	wait, suspend	10,861
13.	/ʔànúyâat/	allow	10,860
14.	/sɔɔn/	teach	10,776
15.	/càtkaan/	manage, deal with	10,418
16.	/hăay/	recover, disappear	10,312
17.	/bankháp/	force, execute, steer	10,222
18.	/duulɛɛ/	take care of	10,216
19.	/còp/	graduate, end/finish	10,051

Table 5.1 Selected verbs, their meanings, and their frequency of occurrence

We have seen from the concordance analysis in Chapter 4 that there are instances where verbs are nominalised (see section 4.4.2.1). Therefore, in this analysis, I will only look at instances where the verb functions as the verb of a clause and exclude instances where the verb is nominalised. That is, I will exclude instances where a verb forms part of a compound noun. For example, in the analysis of /còp/ ‘end’, I excluded /tɔɔn còp/, which is a compound noun meaning ‘ending’. I will also exclude instances where a verb is preceded by the grammatical particles creating a noun, /kaan/, /khwaam/ and /phûu/, such as /kaan pràkàat/ ‘announcement’, /kaan càtkaan/ ‘management’, /khwaam níyom/ ‘popularity’, and /phûu ʔaasăy/ ‘inhabitant’.

I will also exclude instances that involve tokenisation errors. For example, the concordance for /càtkaan/ ‘manage, deal with’ also includes instances where /càtkaan/ is actually a result of tokenisation error. Here, /càt/ is a verb, meaning ‘arrange’, that is followed by a separate unit /kaan/, which is a grammatical particle that nominalises what follows it. Instances involving such tokenisation errors and other similar cases will be excluded. There are also repeated concordance lines in the data¹¹. In cases of duplicated lines, only one will be included in the analysis.

5.5 Results

5.5.1 /pràkàat/ ‘announce, declare’

/pràkàat/ is most frequently used as a transitive verb with an explicit object. There are only 16 instances where the object of /pràkàat/ is not spelt out, although it can be inferred from context what is announced. This applies across all four genres.

The object of /pràkàat/ may precede or follow /pràkàat/, although in most cases it follows. The object that precedes /pràkàat/ can be a noun, or a clause. The object that follows /pràkàat/ can be a noun, a clause, or a verb.

Most of the nouns that function as the object of /pràkàat/ are abstract nouns, many of which are preceded by /kaan/ or /khwaam/, the grammatical particles that create an abstract noun. In terms of semantics, the object noun has a semantic preference for rules and principles, expressed by words such as /kòtʔayyakaansùk/ ‘martial law’, /náyoobaay/ ‘policy’, and /khâaniyom/ ‘value’. It also has a preference for critical situations, expressed by words such as /sǒŋkhraam/ ‘war’, and /sàthāanākaan chùkchǎn/

¹¹ Such repetitions result from double-inclusion of a text in the TNC by error.

‘emergency situation’. These preferences apply across all four genres. In *fiction*, *newspaper*, and *non-academic*, the object noun also has a semantic preference for results (of something), expressed by words such as /phǒn kaansòp/ and /phǒn sòp/ ‘exam result’, and /phǒn pràkòpkaan/ ‘business profit’. In *fiction*, there also appear three cases where the object noun refers to characteristics, including /khwaam thánontua/ ‘self-importance’, and /khwaam dètđiaw/ and /khwaam mûnđmûn/ ‘determination’.

A clause that functions as the object of /pràkàat/ may come before or after /pràkàat/. It may be preceded by the complementizer /wâa/. However, when preceded by /wâa/, it can only follow /pràkàat/. There are 49 instances where the object of /pràkàat/ is a clause in *fiction*, 28 in *newspaper*, 25 in *non-academic*, and 10 in *academic*.

Example 5.1 (*non-academic*)

khǎw	khəy	pràkàat	wâa	thâa	khay	mii	râaŋ
3SG	used.to	announce	COMP	if	REL	have	quarrel
kàp	mêe	kô	mǎan	mii	râaŋ	kàp	màm
with	mom	LP	similar	have	quarrel	with	Mam

‘He once announced that if anyone has a quarrel with mom, it is like they have a quarrel with him.’

/pràkàat/ can alternatively be followed by a verb. This pattern forms a serial verb structure (see section 1.3.3.2). In most such cases, /pràkàat/ is immediately followed by the verb complement. There are seven cases where the verb complement after /pràkàat/ is preceded by the challengeability marker /cà/.

Example 5.2 (*newspaper*)

naay	wɔŋsàk	pràkàat	cà	róŋriian
Mr	Wongsak	announce	CM	sue
naay	sòmkhít			
Mr	Somkit			

‘Mr Wongsak announced that he would sue Mr Somkit.’

The verb complements that follow /pràkàat/ are various and cannot be easily grouped into categories in terms of meaning, and they do not show a tendency towards positive or negative polarity. However, in *academic*, *newspaper*, and *non-academic*, one verb occurs frequently, /cháy/ ‘use’. /pràkàat cháy/, which means ‘announce the use of’, is frequently followed by words expressing rules and regulations, i.e. /kòtʔayyakaansùk/ ‘martial law’, /pràmuan kòtmäay/ ‘codes of law’, and /phrârâatchábanyàt/ ‘act of a legislature’. Even though the sequence /pràkàat cháy/ followed by words expressing rules and regulations occurs frequently, it is not an extended unit of meaning in Sinclair’s sense, as the meaning the sequence conveys is no more than the combination of the meanings of each element of the sequence. This is, rather, a case of semantic preference.

/pràkàat/ may alternatively be followed by a clause preceded by the causative marker /hây/. This creates a causative serialisation (see section 1.3.3.2). Overall, there are 30 instances of /pràkàat hây/ across all four genres.

Example 5.3 (*newspaper*)

hunsen	pràkàat	hây	tháhään	kamphuuchaa
Hun Sen	announce	CAUS	solider	Cambodia
thriamphróom	hàak	mii	kaan	lûanlám
get.ready	if	have	NMLZ	invade

‘Hun Sen announced that the Cambodian soldiers must be ready if there is an invasion.’

Example 5.4 (*newspaper*)

wanthîi	31	thùlaakhom	rátthàbaan	pràkàat	hây
date	31	October	government	declare	CAUS
pen	wanyüt	râatchákaan			
COP	holiday	government			

‘The government declares the 31st of October to be the national holiday.’

In Examples 5.3 and 5.4, the subject of the main clause makes the action in the subordinate clause happen. Thus, it might be argued that the sequence containing /pràkàat hây/ is an extended unit of meaning, which can be laid out as follows:

[person with authority] /**pràkàat hây**/ [person/thing] [verb] ([object/adverb])

This unit of meaning has /pràkàat hây/ as the core. The unit expresses a type of announcement in which a person, particularly one with authority, reports that he/she will make something happen, which is thus its pragmatic function or semantic prosody.

I found that in general the subject of /pràkàat/ tends to be a person or people. There are only 20 cases where the subject is not human, such as /kaanlûaktâj/ ‘election’, and /sěej ʔòon ʔùn/ ‘warm sunlight’. It is evident from the data that in *academic*, the subject of /pràkàat/ is likely to be a person or group of people with authority, particularly one involved with politics, such as a president, prime minister, political party leader, king, government, or cabinet. As in *academic*, the subject of /pràkàat/ in *newspaper* tends to be people or a group of people with authority. However, there are also 30 instances where the subject of /pràkàat/ is a corporation or organisation, such as Toyota Motor, Tesco Lotus, UNESCO, or the Ministry of Public Health, a type that is absent in *academic*. In *fiction*, by contrast, the subject of /pràkàat/ tends to be ordinary people, in 43 cases characters in the story. There are also 26 instances of pronoun subjects, such as /chăn/, /phôm/, or /raw/ (first person pronoun), and /lòon/ or /man/ (third person pronoun). There also exist cases where the subject is a person’s body part or physical action, such as /duañtaa/ ‘eye’, /nâataa/ ‘face’, /yím/ ‘smile’ or /thâathaaj/ ‘gesture’. In

non-academic, on the other hand, there is a mixture of the types of subjects common in *academic*, *newspaper*, and *fiction*.

We have seen that /pràkàat/ is most frequently used as a single-word unit of meaning. However, the verb arguably forms part of an extended unit of meaning when it is followed by the causative marker /hây/, where the unit has a pragmatic function that seems to be distinct from the literal meaning of the core of the unit. When /pràkàat/ is used as a single-word unit of meaning, it still requires colligations and semantic preferences that are typical of the class of words it belongs to. As a reporting verb, /pràkàat/ colligates with a subject which in turn has a semantic preference to be human, although there appears to be some difference in the type of people across the four genres. It also requires an object, which can be a noun, a verb, or a clause. The object noun generally has a semantic preference for abstractness across all four genres.

5.5.2 /càp/ ‘touch, arrest, etc.’

/càp/ has several meanings. In my analysis, I found that the patterns around /càp/ vary according to the meaning the verb conveys.

The most common meaning of /càp/ is ‘touch’. This meaning appears first in the list of meanings of /càp/ given by the *Royal Institute’s Dictionary*, the official dictionary of Thailand. With this meaning, /càp/ is usually immediately followed by an object noun. There are also 14 instances where the object noun of /càp/ comes before the verb or is implicit. In terms of semantics, these object nouns refer to various things, all of them concrete nouns. The most frequent object noun is /mũũ/ ‘hand’. /càp mũũ/ is an interesting case. It literally means ‘touch the hand (of someone)’ or ‘shake hands (with

someone)’. Figuratively, it means ‘cooperate’. This figurative meaning is prevalent in *newspaper* (17 instances), as shown in Example 5.5. There is only one instance in each of *academic* and *non-academic*, and none in *fiction*.

Example 5.5 (*newspaper*)

fàay	rátthàbaan	càp	muu	kàp	fàay	kháan
side	government	touch	hand	with	side	oppose
lòj	phúunthîi	chûaylǎa			pràchaachon	
descend	area	help			people	

‘The government cooperated with the opposition in helping people in the (disaster) area.’

It might thus be argued that /càp muu/ forms part of an extended unit of meaning, which can be laid out as follows:

[person/organisation] /càp muu/ (/kàp/) [person/organisation] [verb] ([object/adverb])

This extended unit of meaning has /càp muu/ as the core of the unit. It has the pragmatic function of expressing a report of cooperation, which is thus its semantic prosody.

/càp/ can also mean ‘arrest (by the police)’ or ‘catch (people)’. With this meaning, /càp/ appears in both active and passive structures. The active and passive structures can be laid out as follows:

Active structure

[person who arrests] /càp/ (/tua/) [person who is arrested]

Passive structure

[person who is arrested] /thùuk/ [person who arrests] /càp/ (/tua/)
/doon/

/tua/ means ‘body’. It is optional in both structures. /thùuk/ and /doon/ are two of Thai’s passive auxiliaries. Unlike other passive auxiliaries, such as /dâyráp/, they usually indicate adverse situations (Iwasaki and Ingkaphirom 2005: 313). It is evident from the data that here /càp/ colligates specifically with the passive auxiliaries /thùuk/ and /doon/, and not with other passive auxiliaries.

Even though these two patterns occur frequently, I would argue that they are not an extended unit of meaning in Sinclair’s sense. This is because the overall pragmatic function of the unit is not distinct from the combination of the meanings of the elements of the unit. If /tua/ were compulsory, I would argue that /càp tua/ is the core of an extended unit of meaning with the pragmatic function of expressing that someone is arrested. But the evidence shows that /tua/ is optional, and whether or not /tua/ is present, /càp/ means ‘arrest’. Therefore, the two patterns above should be considered as the common patterns of the use of the metaphorical sense of /càp/ meaning ‘arrest’ rather than an extended unit of meaning in the Sinclairian sense.

There are 58 instances across all four genres where /càp/ is used with animals as direct object, mostly in the sense of animals being caught for food. Here, /càp/ mostly appears in the active structure.

Metaphorical meaning of /càp/ other than ‘arrest’ can also be observed in the data. In 63 instances, the meaning /càp/ conveys depends on the (abstract) noun that follows /càp/. For example, /càp khwaamrúusùk/ means ‘sense a feeling’, whereas /càp pràden/ means ‘grasp a point’. Unlike /càp mæu/, I would argue that /càp khwaamrúusùk/ and /càp pràden/ are merely collocations, not the cores of the extended units of meaning in the Sinclairian sense. Even though in /càp khwaamrúusùk/ and /càp pràden/, /càp/ is used

metaphorically, the overall meaning of the unit is still a composition of the meanings of the individual elements of the unit. This is different from /càp mɯɯ/, where the pragmatic function of expressing cooperation is totally distinct from the compositional meaning of ‘touching the hand’. In other words, even though both /càp/ and /mɯɯ/ are used in a metaphorical sense, the combination of the metaphorical sense of /càp/ and /mɯɯ/ does not add up to the meaning of ‘cooperation’. This is not true for either /càp khwaamrúusùk/ or /càp pràden/.

/càp/ also occurs in contexts of causation, as shown in Example 5.6.

Example 5.6 (*fiction*)

wan	sàdɛɛŋ	phîi	cà	càp	phûak	thəə	tɛɛŋtua
day	perform	1SG	CM	touch	group	2SG	get.dressed
hây	ʔék						
CAUS	sexy						

‘On the performance day, I will make you guys put on a sexy outfit.’

With this meaning, /càp/ is arguably the core of an extended unit of meaning. The meaning expressed here has far less to do with the literal meaning of /càp/. Rather, when used in this periphrastic causative construction, /càp/ has a distinct pragmatic function of expressing someone making someone else do something, which is its semantic prosody. The unit can be laid out as follows:

[person] /càp/ [another person] [verb] ([object/adverb])

The last frequent pattern has /càp/ immediately followed by /dây/, the potential auxiliary (Iwasaki and Ingkaphirom 2005: 349). In this context, /càp dâŋ/, literally ‘touch to be able to’, means ‘discover a (bad) secret’. It occurs in both active and passive structures, which can be laid out as follows:

Active structure

[person] /càp dây/ (/wâa/) ([secret/bad thing that has been hidden])

Passive structure

[person] /thùuk/ [another person] /càp dây/ (/wâa/) ([secret/bad thing that has been hidden])
/doon/

Example 5.7 (*fiction*)

ʔèp	ʔaw	hăyplaaará	pay	thaa	sǐi	thooŋ
sneakily	take	pot.of.pickled.fish	ASP	paint	colour	golden
lòk	kεε	léew ná	yaŋ	ʔùtsàa	càp	dây
deceive	3SG	ASP PP	still	try	touch	POT
ʔiik	wâa	man	kée			
more	COMP	3SG	fake			

‘(I) painted the pot of pickled fish with golden colour to deceive him, but he discovered that it was a fake one.’

Out of 23 instances, there are 17 where the secret or bad thing that has been hidden is not preceded by /wâa/, but rather is expressed in the preceding clause or discourse. These two patterns are arguably two forms of an extended unit of meaning, sine they have the same semantic prosody, that is, a (bad) secret being discovered. This pragmatic meaning is not evident from any of the words of the unit but is rather spread across the overall sequence.

We have seen that when /càp/ means ‘touch’ and ‘arrest or catch someone’, it is a single-word unit. This is the most frequent use of /càp/. Even though in this context, /càp/ does not participate in any clear fixed pattern, it requires colligations and semantic preferences that are typical of the word class it belongs to. That is, as a transitive verb encoding a physical action, it requires a human subject and a concrete object for the literal, concrete meaning of ‘touch’ and a human object for the more abstract or metaphorical meaning of ‘arrest’. /càp/ also has colligates of its own (the passive

auxiliaries) that are *not* found for *all* verbs of this type. That said, /càp/ has also been found to form part of three extended units of meaning with clear pragmatic functions distinct from the meaning of the core of the units. There seems to be no clear variation in pattern across genres, as all the patterns discussed, except /càp m̩m̩/ with figurative meaning, appear in all four genres. /càp m̩m̩/ occurs frequently in *newspaper*, once in each of *academic* and *non-academic*, but does not appear in *fiction*. /càp dây/ occurs in all four genres, but most frequently in *fiction*.

5.5.3 /klaaypen/ ‘become’

/klaaypen/ means ‘become’. The dominant pattern of /klaaypen/ across all four genres is as follows:

[subject] /klaaypen/ [complement]

The subject of /klaaypen/ is mostly a noun. There are also 75 instances of pronouns and 71 instances of clauses. In terms of semantics, the subject can refer to various things, both concrete and abstract. Many of the abstract nouns are formed with /kaan/ or /khwaam/, the grammatical particles creating a noun. Interestingly, there are a lot of instances of human subjects across all four genres, especially in *fiction*, where there are as many as 84 human subjects. There are 44 instances of human subjects in *newspaper*, 42 in *academic*, and 39 in *non-academic*.

Before investigating /klaaypen/, I had the impression that it would be interesting to look at whether the changes it refers to are positively or negatively evaluated. Therefore, I paid particular attention to the evaluation of the becoming, that is, the complement of the pattern. The complement is mostly a noun, which can be concrete or

abstract. In all four genres, half or more nouns are semantically neutral. But the number of negative nouns is larger than that of positive nouns in all four genres. In fact, in *fiction* and *non-academic*, the proportion of negative nouns to positive nouns is more than double (72 to 29 in *fiction* and 61 to 28 in *non-academic*). In *academic* and *newspaper*, on the other hand, the ratio of negative to positive nouns is relatively less (55 to 30 in *academic* and 54 to 42 in *newspaper*). One frequent negative noun is /panhãa/ ‘problem’. It occurs six times in each of *academic* and *newspaper* and three times in *non-academic*.

Looking at instances where the subject is a human being, I found that in *fiction*, *newspaper*, and *non-academic*, the changes are mostly negative, followed by neutral, with positive changes being least common. In *academic*, on the other hand, the changes are mostly neutral. Still, there are more instances of negative changes than instances of positive changes.

In this pattern, /klaaypen/ is arguably used as a single-word unit of meaning. The meaning the sequence expresses is simply the combination of the literal meanings of the individual words in the pattern, which is ‘something becomes something else’. That said, it is evident from the data, especially in *fiction* and *non-academic*, that the complement nouns tend to be semantically negative more often than positive, although the majority are semantically neutral. Despite this tendency, such sequences do not qualify as extended units of meaning in the Sinclairian sense, because, as earlier argued, they do not exhibit a pragmatic meaning beyond the combination of the meanings of the parts.

5.5.4 /hǎn/ ‘turn’

/hǎn/ literally means ‘change position’. With this meaning, it can be used both transitively and intransitively. /hǎn/ can also be used with a more metaphorical, abstract meaning. When used metaphorically, /hǎn/ forms part of an extended unit of meaning in Sinclair’s sense. I will consider the distribution of literal and metaphorical use across the genres.

In *academic* and *newspaper*, /hǎn/ is predominantly used in its metaphorical, abstract sense. In this use, /hǎn/ is almost always followed by /klàp maa/, /klàp pay/, /maa/, or /pay/. /klàp/ means ‘return’. /maa/ and /pay/ are directional auxiliary verbs (Iwasaki and Ingkaphirom 2005: 18). Literally, /maa/ means ‘come’, and /pay/ means ‘go’. These sequences are in turn followed by a verb, as shown below. Overall, the sequence forms a serial verb structure, as follows:

[person]	/hǎn/	/klàp maa/	[verb]	([object/adverb])
		/klàp pay/		
		/maa/		
		/pay/		

Example 5.8 (*newspaper*)

khànanii	khon	thay	tôŋ	hǎn	klàp	maa
now	people	Thai	must	turn	return	DIR
chûay	kan	khúm	khroŋ	thîidin	phûa	kàsèettràkam
help	REC	protect		land	for	agriculture

‘Now Thai people have to start helping to protect agricultural land.’

This fixed sequence is arguably an extended unit of meaning in Sinclair’s sense, with /hǎn/ as the core. Overall, the sequence has an inceptive meaning of ‘someone starting to do something’. This inceptive pragmatic meaning, which is not evident from any word in the sequence, but rather spread across the whole unit, is therefore the unit’s semantic prosody.

There are also 9 instances of literal intransitive use of /hǎn/ in *academic* and 44 in *newspaper*. In this use, /han/ is also followed by /klàp maa/, /klàp pay/, /maa/, or /pay/, which is also followed by a verb. However, in this context, the meaning this pattern expresses is ‘someone turns (their body) to do something’. The directional auxiliary verbs here serve their normal grammatical function, and do not combine with /hǎn/ to express a distinct pragmatic function.

There are ten instances of literal transitive use of /hǎn/ in these two genres. In this pattern, /hǎn/ takes a direct object (the thing turned), and the subject is not restricted to human beings. There is one instance in *academic* where the subject is /lòok/ ‘earth’.

In *fiction*, on the other hand, /hǎn/ is mostly used in its literal sense, with the meaning of ‘someone turning (their body) to do something’. As in *academic* and *newspaper*, /hǎn/ is mostly followed by /klàp maa/, /klàp pay/, /maa/, or /pay/, and then a verb. The verbs that appear frequently in this pattern are /mòɔŋ/ ‘look’, /bòok/ ‘tell’, /phûut/ ‘talk’, /thǎam/ ‘ask’, and /yím/ ‘smile’.

Example 5.9 (*fiction*)

thəə	hǎn	maa	yím	hây	chǎn
3SG	turn	DIR	smile	DAT	1SG
‘She turned to smile at me.’					

There are also six instances of transitive /hǎn/ used literally, as well as eight instances of the metaphorical use.

In *non-academic*, like in *academic* and *newspaper*, there are more instances of metaphorical than of literal use, although the difference in the proportion is not as big as in *academic* and *newspaper*. Moreover, in *non-academic*, there are also 17 instances

where the subject of intransitive /hǎn/ used literally is a non-human being, such as /lôok/ ‘earth’ and /khûalôok nǎa/ ‘north pole’.

In sum, when used literally, /hǎn/ is arguably a single-word unit of meaning. The meanings of the patterns where /hǎn/ occurs are merely the combination of the meanings of the individual elements of the sequences. In these combinations, /hǎn/ has requirements for colligations and semantic preferences that are typical of the word class it belongs to. That is, as a verb of motion, it requires a subject, which tends to be human, and a direct object when used transitively. However, when used metaphorically, /hǎn/ is the core of an extended unit of meaning expressing the pragmatic meaning of ‘someone starting to do something’. In terms of variation across genres, we have seen that the metaphorical use is prevalent in *academic* and *newspaper*, whereas the literal use dominates in *fiction*. There is not a big difference in the proportion in *non-academic*.

5.5.5 /ʔaasǎy/ ‘live, rely on, refer to’

/ʔaasǎy/ has three different meanings. Literally, it means ‘live’ (in the sense of ‘dwell’). Metaphorically, it means ‘rely on’, or ‘refer to’. The patterns that /ʔaasǎy/ exhibits vary according to the meaning it conveys, however.

When /ʔaasǎy/ means ‘live’, it is used intransitively. With this meaning, /ʔaasǎy/ is frequently followed by /yùu/, the marker indicating continuous aspect (Iwasaki and Ingkaphirom 2005: 152-153). Here, /yùu/ is a colligate. /ʔaasǎy yùu/ is frequently modified by a prepositional phrase starting with /nay/ ‘in’ or /thîi/ ‘at’. The prepositional phrase indicates a location where the subject, which has the semantic preference to be human or animate, lives. There are also 14 cases where /ʔaasǎy yùu/ is immediately

followed by a noun indicating a location without any preposition in between. In nine instances, /ʔaasǎy yùu/ is followed by the preposition /kàp/ ‘with’, indicating a person whom the subject lives with. In addition to the sequence where /ʔaasǎy/ is followed by /yùu/, there are 14 instances where /ʔaasǎy/ is immediately followed by a noun indicating a location, or the preposition /kàp/ followed by a human object. Even though the sequence /ʔaasǎy yùu/ occurs frequently, it does not seem to constitute extended units of meaning. The meaning the sequence expresses is no more than the combination of the meaning of /ʔaasǎy/ and the function of /yùu/ of indicating continuous aspect. There is not a pragmatic function beyond that.

When /ʔaasǎy/ means ‘rely on’ or ‘refer to’, it is almost always used transitively and is followed by a noun as direct object. The object nouns can be concrete or abstract, and many are marked by /kaan/ or /khwaam/, the grammatical particles creating a noun. In *academic*, *newspaper*, and *non-academic*, the abstract nouns outnumber the concrete nouns. Especially in *academic*, the abstract meaning is more than five times more frequent than the concrete meaning (85 to 16). In *fiction*, by contrast, the concrete meaning is more common.

With the meaning of ‘rely on’, /ʔaasǎy/ can alternatively be used intransitively. Here, /ʔaasǎy/ may be preceded by /phûŋphaa/ or /ʔiŋ/, both of which also mean ‘rely on’. /phûŋphaa ʔaasǎy/ and /ʔiŋ ʔaasǎy/ represent a type of semantic reduplication common in Thai (Iwasaki and Ingkaphirom 2005: 36). /(phûŋphaa/ʔiŋ) ʔaasǎy/ is always followed by the reciprocal marker /kan/ or /kan lé kan/, meaning ‘each other’. The subordinator /sûŋ/ may occur between /(phûŋphaa/ʔiŋ) ʔaasǎy/ and /kan (lé kan)/. There are ten instances of this intransitive use in *non-academic*, two in *academic* and one in

fiction. It seems obvious from the data that /(phânpħaa) ʔaasăy (sâŋ) kan (lé kan)/ and /(ʔiŋ) ʔaasăy (sâŋ) kan (lé kan)/ constitute a fixed phraseology. /phânpħaa/ and /ʔiŋ/ are collocates, and /sâŋ/ and /kan (lé kan)/ are colligates. Nevertheless, examining the context, I did not find any further commonality, except the fact that the subject of the verb phrase is two or more people or things. In addition, the phrase does not exhibit other motivations for its use, except the expression of the literal meaning of the phrase of ‘depending on each other’. Therefore, I would argue that although /(phânpħaa) ʔaasăy (sâŋ) kan (lé kan)/ and /(ʔiŋ) ʔaasăy (sâŋ) kan (lé kan)/ occur repeatedly and have their own specific colligates and collocates, they are merely a fixed phraseology, rather than an extended unit of meaning in the Sinclairian sense. This is because they do not exhibit the defining feature, specifically a clear pragmatic function separate from the combined meanings of the parts, of an extended unit of meaning.

In sum, /ʔaasăy/ is arguably a single-word unit of meaning, whether used literally or metaphorically. It does not seem to form part of any clear extended units of meaning. However, it has general requirements for colligations and semantic preferences that are typical of the class of words it belongs to. When it means ‘live’, it colligates with a subject which in turn has the semantic preference to be animate. It also has a specific colligation with the continuous aspect. When it means ‘rely on’ or ‘refer to’ and is used transitively, it also colligates with a subject, although in this context, the subject can also be inanimate. It also has a requirement for a noun as direct object, which can be either concrete or abstract. When intransitive in this sense, it colligates with the reciprocal marker. There does not seem to be obvious variation across genres, except that uniquely in *fiction*, when /ʔaasăy/ means ‘rely on’ or ‘refer to’, there appear to be more concrete

direct objects than abstract direct objects. In addition, in all genres, except in *newspaper*, the instances of /ʔaasǎy/ meaning ‘rely on’ or ‘refer to’ outnumber the instances of /ʔaasǎy/ meaning ‘live’.

5.5.6 /yɔɔm/ ‘resolve, agree’

/yɔɔm/ means ‘resolve’ or ‘agree’. When it means ‘resolve’, /yɔɔm/ most frequently appears as the first verb in a serial verb structure. The verb that follows /yɔɔm/ is mostly semantically neutral; this applies across all four genres, albeit always with more instances of negative verbs than of positive verbs. One frequent negative verb that follows /yɔɔm/ is /phéε/ ‘lose’. /yɔɔm phéε/, literally ‘resolve to lose’, means ‘give in’. This sequence is arguably a collocation, rather than the core of an extended unit of meaning. The meaning the sequence conveys is not clearly more than the combination of the literal meanings of the elements of the sequence.

/yɔɔm/ is also frequently followed a clause preceded by the causative marker /hây/. The sequence /yɔɔm hây/, which is a causative serialisation like the one discussed in 5.5.1, appears across all four genres and can be laid out as follows:

[person] /yɔɔm hây/ [another person/thing] [verb] ([object/adverb])

This sequence is arguably an extended unit of meaning whose core is /yɔɔm hây/. Overall, the unit expresses the meaning of the subject, which has a semantic preference to be human, allowing someone to do something, or allowing something to happen. This pragmatic meaning is quite distinct from the literal ‘resolve to cause’, and is not evident from any word in the sequence, but rather is spread across the whole unit; it is thus the unit’s semantic prosody.

Example 5.10 (*fiction*)

chǎn	mây	chây	khon	thii	cà	yɔɔm	hây	thəə
1SG	NEG	yes	person	SBR	CM	resolve	CAUS	2SG
maa	thamráay	râan̄kaay	day	fii	fii	ròk		
ASP	hurt	body	POT	free	free	PP		

‘I am not a person who allows you to physically attack me without attacking you back.’

When /yɔɔm/ means ‘agree’, it takes a noun, mostly referring to a person, as its direct object. The meaning this pattern conveys is that someone ceases resistance to someone else doing what they prefer. The combination of /yɔɔm/ in this sense and an object is arguably not an extended unit of meaning, because the meaning of the sequence is no more than the combination of the meanings of the elements of the sequence.

/yɔɔm/ is arguably used most frequently as a single-word unit of meaning. With this use, /yɔɔm/ has the colligation and semantic preference that are typical of the word class /yɔɔm/ belongs to. That is, as a verb of cognition, it colligates with a subject which has the semantic preference to be human. It also colligates with a following verb. Even though /yɔɔm/ is mostly used as a single-word unit of meaning, we have also seen cases where /yɔɔm/ arguably forms part of an extended unit of meaning, namely the unit consisting of /yɔɔm hây/ as the core. This extended unit of meaning has a clear pragmatic function of permission, which is distinct from the literal meaning of either element of the core of the unit. There does not seem to be any obvious variation across genres.

5.5.7 /yók/ ‘lift, refer, move, give, dismiss, draft (a document)’

/yók/ has several meanings. Literally, it means ‘lift’. This meaning appears first in the list of meanings of /yók/ given by the *Royal Institute’s Dictionary*. /yók/ also has

other more metaphorical, abstract meanings, including ‘refer’, ‘move’, ‘give’, ‘dismiss’, and ‘draft (a document)’.

When /yók/ means ‘lift’, /yók/ is often immediately followed by a noun as direct object. There are also 66 instances where a noun functioning as an object of /yók/ comes before the verb or where the object is implicit and must be inferred from context. The object noun of /yók/ is almost always a concrete item. When the direct object of /yók/ comes after the verb, it may in turn be followed by a directional auxiliary verb, such as /khûn/ ‘ascend’, /loŋ/ ‘descend’, or /ʔòk/ ‘exit’. This sequence of /yók/ + direct object + directional auxiliary verb creates a resultative serialisation (see section 1.3.3.2), as shown in Example 5.11.

Example 5.11 (*fiction*)

thəə	yók	krápǎw	khûn	phûa	cà	pəət
3SG	lift	bag	DIR	for	CM	open
bon	tiaŋ					
on	bed					

‘She lifted the bag up, so she could open it on the bed.’

In Example 5.11, the serial verb sequence /(thəə) yók krápǎw khûn/ consists of two events: /(thəə) yók krápǎw/ and /krápǎw khûn/, where the second event is the result of the first event. /krápǎw/ functions as the object of /yók/ and subject of /khûn/.

There are also some instances where the object noun of /yók/ is not concrete, but abstract. There are 15 and 13 instances in *newspaper* and *academic*, but only 5 and 2 in *non-academic* and *fiction*, respectively. The most frequent abstract noun that appears across all four genres is /rádàp/ ‘level’. One example from *academic* is as follows:

Example 5.12 (*academic*)

kàsèettàkoon	nay	phâak	tàwanʔòk	khǒo	hây	rátthàbaan
farmers	in	part	east	beg	CAUS	government
yók	rádàp	raakhaa	ʔôy	hây	sũuŋ	khûn
lift	level	price	sugarcane	CAUS	high	ascend

‘The farmers in the East asked the government to increase the price of sugarcanes.

Even though /yók rádàp/ occurs repeatedly, it is arguably a collocation, rather than the core of an extended unit of meaning. The combination does not exhibit any pragmatic function that is distinct from the combination of the metaphorical meaning of /yók/ ‘lift’ plus the literal meaning of /rádàp/ ‘level’.

‘Lift’ is the most frequent meaning of /yók/ in *fiction*, *newspaper*, and *non-academic*. There are as many as 128 instances of this meaning in *fiction*. In these three genres, there are many more instances where the object noun is concrete than where the object noun is an abstract concept. In *academic*, by contrast, the most frequent meaning of /yók/ is not ‘lift’, but ‘refer’ (92 instances). ‘Lift’ is the second most frequent meaning. Moreover, in *academic*, out of 20 instances where /yók/ means ‘lift’, there are as many as 13 instances where the object noun is an abstract concept.

When /yók/ means ‘refer’, it also takes a direct object, which may come immediately after /yók/, or before the verb. The most frequent object noun of /yók/ with this meaning across all four genres is /tuayàaŋ/ ‘example’. /yók tuayàaŋ/ means ‘give an example’. There are also 34 instances where /yók tuayàaŋ/ is followed by /chên/. /chên/ means ‘for example’ or ‘such as’. /yók tuayàaŋ chên/ altogether also means ‘for example’. Unlike /yók tuayàaŋ/, which behaves similarly to other verb phrases and is thus arguably merely a collocation, I would argue that /yók tuayàaŋ chên/ is the core of an extended unit of meaning, as it has its own specific colligation and semantic prosody.

/yók tuayàaŋ chên/ has the grammatical preference to appear at the beginning of a sentence. It is also always followed by an example of something that has been mentioned in the preceding discourse. Overall, the unit thus has the pragmatic function of linking some topic or phenomenon to an example of that topic or phenomenon, which is its semantic prosody.

[some phenomenon] | /yók tuayàaŋ chên/ [example of phenomenon]

Example 5.13 (*academic*)

khon	pen	rôok	pràsàat	thîi	khít	wâa		
people	COP	disease	neurotic	SBR	think	COMP		
sìŋ	thîi	kèət	khûn	nay	cay	măankàp	kèət	
thing	SBR	happen	ASP	in	mind	similar	happen	
khûn		nôok	tua	thamhây	thôot	tuapræ		
ASP		outside	body	make	blame	variable		
khâaŋnôok	wâa	pen	tôn hèet	khốŋ	panhăa			
outside	COMP	COP	cause	of	problem			
yók	tuayàaŋ	chên		khon	pen	rôok		
lift	example	for.example		people	COP	disease		
pràsàat	mák	khít	wâa	khon	ʔàun	kamlaŋ		
neurotic	often	think	COMP	people	other	ASP		
kròot	tuaʔeeŋ	têe	ciŋ	ciŋ	léew	tua	khăw	
angry	REFL	but	real	real	ASP	body	3SG	
ʔeeŋ	kamlaŋ	kròot	tuaʔeeŋ					
by.oneself	ASP	angry	REFL					

‘People with neurotic disorder think that what happens in their mind is similar to what happens outside their mind. This causes them to blame outside variables. For example, people with neurotic disorder often think that other people are angry at them, but in fact, they are angry at themselves.’

/yók tuayàaŋ chên/ occurs 20 times in *academic*, 12 in *non-academic*, and 2 in *newspaper*. It does not occur at all in *fiction*.

When it means ‘move’, /yók/ also takes a noun as direct object, with a semantic preference for this noun to refer to a group of people. The top object nouns are military, such as /kốŋtháp/ ‘army’, /kamlaŋ/ ‘force’, and /phon/ ‘troop’. /yók kốŋtháp/, /yók

kamlan/, and /yók phon/ all mean ‘make an army go (to a place)’. Even though these sequences occur frequently, they arguably form a collocation, rather than part of any extended units of meaning. The meaning each sequence expresses is no more than the combination of the meanings of the individual elements of the sequence. In addition to these top three collocates, there appear other collocates that are not related to the military, albeit they do refer to groups of people, such as /kɔɔŋ/ ‘production team’ and /kúan/ ‘gang’.

When /yók/ means ‘give’, it is always used as a ditransitive verb. The direct object may come before or after /yók/. The indirect object, which has a semantic preference to be human, always comes after /yók/ and follows the direct object in cases where the direct object appears after /yók/. The indirect object is always preceded by /hây/, literally ‘give’. Here, /hây/ is a dative marker, functioning similarly to English ‘to’.

When /yók/ means ‘dismiss’, it also takes a noun as direct object. Two frequently co-occurring object nouns are /thòot/ ‘penalty’ and /khamróŋ/ ‘(legal) case’. /yók thòot/ means ‘forgive’, and /yók khamróŋ/ means ‘dismiss the case’.

These four different metaphorical meanings of /yók/ occur across all four genres. In *academic*, *newspaper*, and *non-academic*, there are also instances where /yók/ means ‘draft (a document)’, as in /yók râaŋ (rátthàthammánuun)/ ‘draft (the constitution)’. There are 11 instances in *newspaper*, 7 in *non-academic*, and 1 in *academic*. This meaning of /yók/ as ‘draft’ is absent in *fiction*.

I would argue that the combinations of /yók/ and its preferred direct objects mentioned above merely form collocations, rather than extended units of meaning. The pragmatic function each combination expresses is no more than the combination of the

(metaphorical) meanings of /yók/ and of its direct object. They do not exhibit a clear separate semantic prosody.

Whether used literally or metaphorically, /yók/ is arguably a single-word unit of meaning. Whether it means ‘lift’, ‘refer’, ‘move’, ‘give’, ‘dismiss’, or ‘draft (a document)’, the meaning of the sequence where /yók/ occurs is the combination of the meaning of /yók/ and its object(s). /yók/ can be considered part of an extended unit of meaning only when it is used in the fixed sequence /yók tuayàaŋ chên/, which appears most frequently in *academic* but is absent in *fiction*. That said, /yók/ has the requirements for colligation and semantic preference typical of the word class(es) it belongs to. As a transitive verb (of motion), it requires a subject, which in turn has the semantic preference to be human, and an object. When used as a ditransitive verb meaning ‘give’, as well as a human subject, it colligates with a human indirect object. While the meaning of ‘lift’ is most frequent in *fiction*, *newspaper*, and *non-academic*, the meaning of ‘refer’ is the most prevalent in *academic*.

5.5.8 /yũũn/ ‘stand’

/yũũn/ has both a literal and a more metaphorical, abstract meaning. Literally, /yũũn/ means ‘stand’. When used metaphorically, the meaning of /yũũn/ depends on the context and pattern where it occurs. In all four genres, the literal use is much more frequent than the metaphorical use.

When /yũũn/ is used in its literal sense, it most frequently appears in a serial verb construction across all four genres. /yũũn/ may appear in any position in the verb sequence. The verb or verbs that precede /yũũn/ form a sequential serialisation, indicating

two or more actions that happen one after another (Iwasaki and Ingkaphirom 2005: 233). The verb or verbs following /yʉʉn/, on the other hand, form a simultaneous serialisation, where two or more actions happen at the same time (Iwasaki and Ingkaphirom 2005: 236).

Example 5.14 (*newspaper*)

nám tòkcaɣ lúk khûn yʉʉn
 Nam frightened get.up DIR stand
 ‘Nam was frightened, and she stood up.’

In the above example, /yʉʉn/ is preceded by two verbs /tòkcaɣ/ ‘be frightened’ and /lúk/ ‘get up’, the latter of which is modified by /khûn/, the directional auxiliary verb meaning ‘ascend’. The meaning the serial verb sequence expresses is sequential; the subject felt frightened, and then stood up.

Example 5.15 (*fiction*)

raw yʉʉn duu phráʔaathít khûn thîi sàphaanplaa
 1PL stand watch sun rise at pier
 ‘We stood at the pier, watching the sun rise.’

Here, /yʉʉn/ appears as the first verb in the sequence and is followed by /duu/ ‘watch’. The meaning the sequence conveys is simultaneous: the actions of ‘standing’ and ‘watching the sun rise’ are happening at the same time.

/yʉʉn/ also appears outside of the serial verb structure. In this case, /yʉʉn/ is sometimes modified by an adverb, such as /troŋ/ as in /yʉʉn troŋ/ ‘stand straight’ or /kêe kêe kaŋ kaŋ/ as in /yʉʉn kêe kêe kaŋ kaŋ/ ‘stand awkwardly’. It may alternatively be modified by /yùu/, the marker of continuous aspect (Iwasaki and Ingkaphirom 2005: 152-153); this is a clear colligation, as there are 98 instances of /yʉʉn yùu/ across the four

genres. Frequently, /yʉʉn yùu/ further colligates with a prepositional phrase starting with /troŋ/ or /thîi/ ‘at’, /khâaŋ/ ‘beside’, etc., indicating a location where a subject stands.

Example 5.16 (*fiction*)

phûuchaay	thîi	yʉʉn	yùu	khâaŋ	chǎn	tɔɔnníi
man	SBR	stand	ASP	beside	1SG	now
khǎw	pen	khray				
3SG	COP	REL				

‘Who is the man standing beside me?’

When /yʉʉn/ is used metaphorically, the meaning it conveys varies depending on the surrounding context. The metaphorical use is prevalent in *academic* and *newspaper*; there are 18 instances in *academic* and 20 instances in *newspaper*. There are eight instances in *non-academic*, and only two in *fiction*. The pattern inducing a particular metaphorical meaning of /yʉʉn/ that occurs most frequently, and occurs across the four genres, is as follows:

[person]	/yʉʉn/	(/yùu/)	/kàp/	[another person]
			/khâaŋ/	

Example 5.17 (*newspaper*)

thâa	kɔɔŋtháp	thay	mây	dây	yʉʉn	yùu	kàp
if	army	Thai	NEG	POT	stand	ASP	with
pràchaachon	nân	mây	chây	kɔɔŋtháp	thay		
people	DEM	NEG	yes	army	Thai		

‘If the Royal Thai Armed Forces do not stand by Thai people, that is not the Royal Thai Armed Forces.’

The meanings the pattern expresses are (i) standing by someone, in the sense of remaining loyal in a difficult situation, or (ii) taking someone’s side, depending on context. /yùu/ is the continuous aspect marker as noted above. /kàp/ and /khâaŋ/ are prepositions, meaning ‘with’ and ‘beside’, respectively. /yùu/ is compulsory when /kàp/ is selected, but optional with /khâaŋ/. One might argue that this sequence forms an

extended unit of meaning, because it seems to form a rather fixed pattern that has specific colligations (/kàp/ and /khâaŋ/) and a semantic preference (for people). But I would argue that the sequence is merely an example of the standard metaphorical use of /yũũn/. Examining the contexts where the sequence is used, I did not find any commonality in terms of pragmatic motivation for the use of the sequence. The sequence is not used in any specific contexts, and it simply expresses the meaning of ‘someone standing by someone else’, which is the extended metaphorical meaning of /yũũn/. So this repeated sequence does not have its own separate pragmatic meaning, i.e. semantic prosody.

In sum, whether used literally or metaphorically, /yũũn/ is a single-word unit of meaning. Even though it does not form part of any extended units of meaning, it has requirements for colligation and semantic preference typical of the word class it belongs to and a specific colligation with serial verb structures and with the continuous aspect. As a stative intransitive verb, it requires a subject, which in turn has the semantic preference to be an animate being, especially a human being. In terms of variation across genres, we see only a difference in the prevalence of the metaphorical use of the verb.

5.5.9 /khûapkhum/ ‘control, supervise, confine’

/khûapkhum/ means ‘control’, ‘supervise’, or ‘confine (someone)’. The patterns /khûapkhum/ exhibits vary according to the meaning the word expresses.

When /khûapkhum/ means ‘control’ or ‘supervise’, it is frequently immediately followed by a noun as direct object. There are 93 instances where the direct object precedes the verb or is not explicit but must be inferred from context. The object can be either concrete or abstract. This applies across all the four genres. In *academic*,

newspaper, and *non-academic*, 52 of the direct objects are abstract nouns beginning with /kaan/, the grammatical particle that creates a noun. There are only three instances in *fiction*. The nouns following /khûapkhum/ are various in meaning, and cannot be easily grouped into particular semantic categories. Nevertheless, in *fiction*, there is an obvious semantic preference for the object noun to refer to emotion, e.g. /ʔaaron/ ‘mood’, /khwaamrúusàk/ ‘feeling’, and /thoosà/ ‘anger’.

/khûapkhum/ may alternatively be followed by /tuaʔeeŋ/ or /tonʔeeŋ/, both of which are reflexive pronouns. /khûapkhum tuaʔeeŋ/ or /khûapkhum tonʔeeŋ/ means ‘control oneself’. This sequence is prevalent in *fiction* (19 instances). There are not many in *academic*, *newspaper*, and *non-academic* (5, 2, and 7 instances, respectively). Out of these 33 instances, there are 18 instances where the sequence is followed by /dây/ or /mây dâŋ/. /dây/ is a potential auxiliary, literally ‘be able to’ (Iwasaki and Ingkaphirom 2005: 350), and /mây/ is a negator.

Example 5.18 (*fiction*)

chaaynùm	ʔaaron	dii	khûn	khûapkhum	tonʔeeŋ
young.man	mood	good	ASP	control	REFL
dây	mâak	khûn			
POT	much	ASP			

‘The young man was in a better mood, and could control himself better.’

/khûapkhum tuaʔeeŋ (mây) dâŋ/ or /khûapkhum tonʔeeŋ (mây) dâŋ/ seems to be a fixed phraseology. It is evident from the data that /khûapkhum/, when followed by /tuaʔeeŋ/ or /tonʔeeŋ/, has a semantic preference for ability, which is expressed by /dây/, which is in turn a colligate. However, even though the sequence is rather fixed and exhibits specific colligation and semantic preference, I would argue that it is not an extended unit of meaning. This is because the sequence is not used in any specific

contexts, and it does not exhibit a clear pragmatic function, i.e. semantic prosody, that is distinct from the combination of the literal meanings of the elements of the unit.

With the meaning of ‘control’, /khûapkhum/ may also be followed by a verb. Interestingly, most of the verbs that appear after /khûapkhum/ have meanings similar to the meaning of /khûapkhum/, such as /duuleɛ/ ‘take care of’, /kamkàp/ ‘supervise’, or /baŋkháp/ ‘force’. /duuleɛ/ is most frequent and appears across all four genres.

Alternatively, /khûapkhum/ may be followed by /tua/ ‘body’. /khûapkhum tua/ appears in both active and passive structures, as follows.

Active structure

[person who confines] /**khûapkhum tua**/ ([person who is confined])

Passive structure

[person who is confined] /thùuk/ [person who confines] /**khûapkhum tua**/
/doon/

/tua/, which means ‘body’, is compulsory in both active and passive structures. /khûapkhum/ alone means ‘control’. To mean ‘confine’, /khûapkhum/ must be followed by /tua/. /khûapkhum tua/ is thus arguably the core of an extended unit of meaning with the pragmatic function of expressing that someone is confined. This meaning is not evident from any individual part of the element of the unit, i.e. /khûapkhum/ or /tua/, but rather belongs to the combination of /khûapkhum/ and /tua/.

When /khûapkhum tua/ is used in the active structure, its subject has a semantic preference for people with authority, such as /tamrùat/ ‘police’, /câawnâathîi/ ‘officer’, and /rátthàbaan/ ‘government’. When used in the passive structure, it colligates with /thùuk/ and /doon/, two of Thai’s passive auxiliaries. Unlike other such auxiliaries, such as /dâyrap/, these two auxiliaries usually indicate adverse situations (Iwasaki and

Ingkaphirom 2005: 313). This extended unit of meaning is prevalent in *newspaper* (34 instances), whereas it appears only six times in *academic*, twice in *fiction*, and once in *non-academic*.

We saw in section 5.5.2 that /càp/ in the more abstract, metaphorical sense of ‘arrest’ also co-occurs with /tua/, in both active and passive structures. When used in the passive structure, it also specifically colligates with /thùuk/ and /doon/. However, in contrast to the case with /khûapkhum/, in the case with /càp/, /tua/ is optional. Whether or not /tua/ is present, /càp/ can mean ‘arrest’. In other words, the meaning of ‘arrest’ is expressed by /càp/ considered in isolation, rather than by the combination of /càp/ and /tua/. In contrast to /khûapkhum tua/, the case for considering /càp tua/ to be the core of an extended unit of meaning is much weaker.

When it means ‘control’ or ‘supervise’, /khûapkhum/ is arguably a single-word unit of meaning. It has the colligations and semantic preferences that are generally required by the word class it belongs to. That is, as a transitive verb, it requires a subject, which in turn has the semantic preference to be a human being, and requires a direct object. It also has some specific collocations and colligations. That said, /khûapkhum/ also seems to form part of the core of an extended unit of meaning /khûapkhum tua/. We have seen some variation across genres. The specific colligation where /khûapkhum/ is followed by /tuaʔeeŋ/ or /tonʔeeŋ/ is prevalent only in *fiction*, whereas the extended unit of meaning whose core is /khûapkhum tua/ is frequent only in *newspaper*.

5.5.10 /níyom/ ‘like’

/níyom/ displays two major patterns. The first pattern has /níyom/ followed by a direct object, which can be either a concrete or an abstract noun. The majority of these nouns are semantically neutral. The second pattern has /níyom/ as the first verb in a serial verb construction where /níyom/ is followed by a verb complement. In all four genres, the verb-complement pattern outnumbers the noun-object pattern. Like the object nouns, almost all of the verbs after /níyom/ are neutral in meaning. There are five negative verbs in *fiction* and one in each of *academic*, *newspaper*, and *non-academic*. There are five and two positive verbs in *academic* and *newspaper* respectively. The most frequent verb complement across all four genres is /cháy/ ‘use’. /níyom cháy/ means ‘like to use’, as shown in Example 5.19.

Example 5.19 (*newspaper*)

phúakraw	sùanyày	cà	níyom cháy	bòorikaan
1PL	majority	CM	like use	service
khǒŋ	yuuthúp			
of	YouTube			

‘Most of us like to use the YouTube service.’

/níyom/ is arguably a single-word unit of meaning. The meaning the two major patterns of /níyom/ conveys is merely the combination of the meanings of the individual elements of the sequence, that is, a person liking something or liking to do something. This is different from the previously discussed /chôp/ (see section 4.4.2.3). Both /níyom/ and /chôp/ mean ‘like’. They also exhibit the same major patterns: they both take a noun or verb complement as direct object. Nevertheless, unlike /chôp/, /níyom/ does not form part of an extended unit of meaning. It does not pragmatically express the notion of a (bad) habit when used in a serial verb construction, as /chôp/ does; there is no evidence

in the concordance that suggests that what is liked is a habit. However, although /níyom/ does not form part of any extended units of meaning, it does have requirements for colligation and semantic preference common to the word class it belongs to. That is, as a transitive verb of cognition, it requires a subject, which has the semantic preference to be a human being, and requires an object, which can be either a noun or verb.

5.5.11 /waan/ ‘put, be placed, set, walk’

/waan/ has several meanings. Literally, it means ‘put’. It can also be used to indicate the position of the subject, with no implication of motion. Used thus, /waan/ is equivalent to ‘be placed’. It also has other more metaphorical, abstract meanings, such as ‘set’ as in ‘set the rules’ or ‘set a watch’ (in the military sense). The meaning of ‘set a watch’ is not prevalent; it occurs only two times in *newspaper*. /waan/ can also mean ‘walk’. This use is specific to poetry, and in my concordances it appears only once in *fiction*.

When /waan/ means ‘put’, it takes a noun as direct object, which may come before or after /waan/. The object is always a concrete noun. In cases where the object comes after /waan/, the object may in turn be followed by /loŋ/, the directional auxiliary verb meaning ‘ascend’. This creates a resultative serial verb construction, similar to the case of /yók/, discussed earlier. The sequence of /waan/ + direct object + /loŋ/ may be followed by a prepositional phrase indicating where the object is put. The sequence of /waan/ + direct object may alternatively be immediately followed by a prepositional phrase without /loŋ/ in between. The subject of /waan/ with this meaning is always a human being.

When it means ‘put’, /waan/ is sometimes followed by a verb. Two verbs that appear after /waan/ in the data are /khăay/ and /camnàay/, both of which mean ‘sell’. /waan khăay/ and /waan camnàay/ mean ‘put on sale’. These two sequences are arguably collocations, rather than the core of any extended units of meaning. The meaning the sequence conveys is no more than the combination of the literal meanings of the words of the sequence. /waan khăay/ and /waan camnàay/ appear only in *fiction*, *newspaper*, and *non-academic*, not in *academic*.

When /waan/ means ‘be placed’, it is frequently followed by /yùu/, the continuous aspect marker (Iwasaki and Ingkaphirom 2005: 152-153). /waan yùu/ is in turn frequently followed by a prepositional phrase indicating where the subject is. There are four instances where /waan/ is immediately followed by a prepositional phrase without /yùu/ in between. The subject of /waan/ with this use is always a concrete noun.

One frequent pattern of /waan/ meaning ‘be placed’ in a metaphorical rather than literal sense is as follows:

/waan/ /yùu/ /bon/ [concept]

This pattern is prevalent in *academic*. In this context, the subject of /waan/ is always an abstract noun. The meaning the sequence expresses is something being based on some concept, as shown in Example 5.20.

Example 5.20 (*academic*)

khrooŋsâaŋ	săŋkhom	sàkdinaa	thîi	waan	yùu	bon
structure	society	feudal	SBR	put	ASP	on
rábòp	ʔùppàthăm					
system	patronise					

‘the structure of feudalism that is based on a system of patronage’

The above sequence is rather fixed. It has its own colligations, the aspect marker /yùu/ and the preposition /bon/, and its own semantic preference (for a concept). That said, I would argue that the sequence does not form an extended unit of meaning. The pragmatic function of the unit is not distinct from the combination of the meanings of the metaphorical sense of /waan/ as ‘based on’ and the other elements of the sequence. In other words, as /waan/ also has the metaphorical sense of ‘based on’ in other contexts, this sequence is better seen as a fixed phraseology rather than an extended unit of meaning.

When it means ‘put’ in a metaphorical sense, the meaning that /waan/ conveys depends on the object that follows it and the context where it occurs. For example, whereas /waan ton penklaan/, literally ‘put oneself unprejudiced’, means ‘be unprejudiced’, /waan màak/, literally ‘put chess’, means ‘plan a winning strategy’. In the former example, /waan/ has not much more meaning than a copula. Unlike /waan ton penklaan/, however, /waan màak/ seems to be an extended unit of meaning, given that it has a pragmatic function that is distinct from the combination of the meanings of /waan/ ‘put’ and /màak/ ‘chess’. That said, there are only two instances of /waan màak/ being used in this sense in my data. Thus, I cannot state with confidence that /waan màak/ is an extended unit of meaning. To argue for this confidently, more examples would be needed.

The use of /waan/ meaning ‘put’ or ‘be placed’ literally is the most frequent use in *fiction*, *newspaper*, and *non-academic*. The corresponding metaphorical senses ‘put’ and ‘be placed’ appear across all four genres, especially in *academic*.

When it means ‘set’, /waŋ/ also takes a noun as direct object, which may precede or follow /waŋ/. Some of the frequently co-occurring object nouns are /lākkaan/ ‘principle’, /lākkeen/ ‘regulation’, /ŋhānkhǎy/ ‘condition’, and /pāwmǎay/ ‘target’. These combinations arguably form collocations. This sense of /waŋ/ is the most frequent use in *academic*. There are 48 examples in *newspaper*, 20 in *non-academic*, and only 5 in *fiction*.

In sum, whether used literally or metaphorically, /waŋ/ is arguably a single-word unit of meaning. Even though it does not seem to form part of any extended units of meaning, it has colligation and semantic preference typical of the word class it belongs to. As a transitive verb (of motion), it colligates with a subject, which has the semantic preference to be human, and with an object. When used intransitively meaning ‘be placed’, by contrast, it colligates with a subject which has the semantic preference to be a concrete object in literal usage, but an abstract concept in metaphorical usage. Although the literal meanings of ‘put’ or ‘be placed’ are more frequent in *fiction*, *newspaper*, and *non-academic*, the metaphorical sense is more frequent in *academic*.

5.5.12 /rɔɔ/ ‘wait, suspend’

/rɔɔ/ has literal and metaphorical meanings. Literally, it means ‘wait’. Metaphorically, it means ‘suspend’ as in ‘suspend a sentence’. The literal use of /rɔɔ/ dominates across all four genres. The metaphorical use is present only in *academic* and *newspaper*.

When /rɔɔ/ means ‘wait’, it most frequently takes a noun or personal pronoun as direct object, which almost always comes after /rɔɔ/. This applies across all four genres.

The object noun can be either concrete or abstract and expresses what the subject, which has a semantic preference to be human, is waiting *for*. The object noun can also be human, especially in *fiction*. There are also 177 instances where the object of /rɔɔ/ is not present, but can be inferred from context. In this situation, /rɔɔ/ may be followed by /yùu/, the continuous aspect marker, which is in turn frequently followed by a prepositional phrase indicating where the subject is waiting; these are colligations. /rɔɔ/ may alternatively be followed by a clause preceded by the conjunctions /con/ or /conkwàa/, both of which mean ‘until’. It may also be followed by a clause preceded by the causative marker /hây/, which in this context also means ‘until’.

Alternatively, /rɔɔ/ meaning ‘wait’ may occur in a serial verb construction, followed by a verb which functions as the complement of /rɔɔ/. Two frequent verbs across all four genres are /duu/ ‘look’ and /ráp/ ‘receive’. /rɔɔ duu/ means ‘wait to see’. The meaning of /rɔɔ ráp/ varies, depending on the object noun that comes after /ráp/ and the wider context. For example, whereas /rɔɔ ráp thun/ means ‘wait to receive a scholarship’, /rɔɔ ráp (person)/ means ‘wait to pick (someone) up’.

When /rɔɔ/ means ‘suspend’, it also most frequently takes a noun as direct object. Two frequent object nouns are /kaan kamnòt thòot/ ‘determination of punishment’ and /kaan loŋthòot/ ‘punishment’. /rɔɔ kaan kamnòt thòot/ means ‘suspend determination of punishment’, and /rɔɔ kaan loŋthòot/ means ‘suspend a punishment’. We also see examples in one serial verb structure, namely with /loŋ ʔaayaa/. /loŋ/ means ‘descend’, and /ʔaayaa/ means ‘punishment’; /rɔɔ loŋ ʔaayaa/ means ‘suspend a sentence’. These three phraseological units are arguably the cores of three separate extended units of meaning. They are used only in legal contexts to express the act of suspending

determination of punishment and the act of suspending a sentence, which is thus the units' semantic prosody. There are four instances of /rɔɔ kaan kamnòt thòot/ and five instances of /rɔɔ kaan lonthòot/ in *academic*. /rɔɔ lon ʔaayaa/ occurs twice in *academic* and three times in *newspaper*.

In sum, when /rɔɔ/ means 'wait', it is arguably always used as a single-word unit of meaning. It has the requirements for colligation and semantic preference typical of this type of verb. That is, as a transitive verb, it requires a subject which has the semantic preference to be human, and an object, although in 177 instances, the object is implicit. This use of /rɔɔ/ meaning 'wait' is prevalent across all four genres. When /rɔɔ/ means 'suspend', by contrast, it forms part of the core of three separate extended units of meaning with a clear semantic prosody. This use of /rɔɔ/ is technical legal language, found only in *academic* and *newspaper*.

5.5.13 /ʔànúyâat/ 'allow'

/ʔànúyâat/ means 'allow'. There are two major patterns of /ʔànúyâat/ across the four genres.

In the first pattern, /ʔànúyâat/ is used in a serial verb construction. It appears as the second verb following /khǒɔ/ 'ask for'. /khǒɔ ʔànúyâat/ means 'ask for permission', and may be followed by a verb indicating the action the subject asks for permission to perform. In cases where there is not such a verb, the action the subject asks for permission to perform can be inferred from context, usually from the preceding discourse. Between /khǒɔ ʔànúyâat/ and the verb, there may be a person whose

permission is sought, which may in turn be preceded by /càak/, the preposition meaning ‘from’.

Example 5.21 (*fiction*)

phǎm khǎo ʔànúyâat phôo rian tǎo pàrinyaathoo
 1SG ask.for allow father study next master’s degree
 ‘I asked for permission from my father to pursue a master’s degree.’

Even though the sequence /khǎo ʔànúyâat/ is fixed and occurs frequently, I would argue that it does not seem to be an extended unit of meaning. This is because the meaning the sequence expresses is the combination of the meanings of each element of the sequence (‘ask for’ plus ‘allow’). There is no pragmatic function that belongs strictly to the combination.

In the second pattern, /ʔànúyâat/ is followed by a clause preceded by the causative marker /hây/. This creates a causative serialisation (see section 1.3.3.2). /ʔànúyâat hây/ may appear in an active or a passive structure. In all genres, the active use is more frequent than the passive use. When it appears in an active structure, the meaning /ʔànúyâat hây/ expresses is ‘allow someone to do something.’ This pattern can be laid out as follows:

[person who allows] /ʔànúyâat hây/ [person who is allowed] [verb] ([object/adverb])

The subject of /ʔànúyâat hây/ across all four genres tends to be a person or organisation who in some way has authority over the person granted permission, albeit with some variation. In *academic* and *newspaper*, the subject of /ʔànúyâat hây/ tends to be a ruling body, such as /sǎan/ ‘court’, /rátthàbaan/ ‘government’, /kòtmǎay/ ‘law’, or /khánárátthàmontrii/ ‘cabinet’. In *fiction*, by contrast, most of the subjects of /ʔànúyâat

hây/ (5 out of 58) are non-official people who are characters in the stories, but who do have authority over another character in the story, such as a father or mother who has authority over their children. In *non-academic*, there is a mixture of the types of subject appearing in the other three genres.

Example 5.22 (*newspaper*)

sămnákhaan	truat	khon	khâw	muan	ʔànúyâat	hây	naay
office	inspect	people	enter	city	allow	CAUS	Mr
roobót	dəonthaan	khâw	prâthêet	thay		dây	
Robert	travel	enter	country	Thailand		POT	

‘The Immigration Bureau allowed Mr Robert to enter Thailand.’

When /ʔànúyâat hây/ is used in a passive structure, it is always preceded by /dâyráp/, a passive auxiliary. Unlike other such auxiliaries, such as /thùuk/ or /doon/, /dâyráp/ usually indicates favourable situations (Iwasaki and Ingkaphirom 2005: 317). Here, the meaning the sequence /dâyráp ʔànúyâat hây/ expresses is ‘(someone) being allowed to do something.’ The passive pattern can be laid out as follows:

[person who is allowed] /dâyráp/ /ʔànúyâat hây/ [verb] ([object/adverb])

Example 5.23 (*non-academic*)

khăw	dâyráp	ʔànúyâat	hây	dəonthaan	pay
3SG	PASS	allow	CAUS	travel	DIR

sâhârát
America
‘He was allowed to travel to America.’

Whether active or passive, /ʔànúyâat hây/ is arguably the core of an extended unit of meaning that has the pragmatic function of expressing permission being granted. The sequence has specific colligations and semantic preferences. When used in an active structure, it has a semantic preference for a subject to be human and to be in authority. When used in a passive structure, it specifically colligates with the passive auxiliary

/dâyráp/. This semantic preference and colligation apply specifically to the sequence /ʔànúyâat hây/, rather than any individual element of the sequence.

There are also 70 instances where /ʔànúyâat/ is used outside these two patterns. In these instances, what is allowed or not allowed is not explicit as there is no complement, but can be inferred from context, usually in the preceding discourse.

Example 5.24 (*non-academic*)

khăw	khəy	khǒ	klàpbâan	dəəm	lăay	khraŋ
3SG	used.to	ask.for	return.home	original	many	time
tèe	càkkràphát	mây	soŋ	ʔànúyâat		
but	king	NEG	HON	allow		

‘He used to ask to return to his hometown many times, but the king didn’t allow.’

In sum, we have seen that /ʔànúyâat/ is frequently used in two fixed patterns, /khǒ ʔànúyâat/ and /ʔànúyâat hây/. In the former pattern, /ʔànúyâat/ is arguably used as a single-word unit of meaning because the meaning the sequence conveys is no more than the combination of the meanings of /khǒ/ and /ʔànúyâat/. Used this way, the sequence colligates with a human subject as well as a verb following /ʔànúyâat/. In the latter pattern, by contrast, /ʔànúyâat/ arguably forms part of the core of the extended unit of meaning. /ʔànúyâat hây/ has specific colligations and semantic preferences as well as a clear semantic prosody of expressing permission being granted. When used in the active, it colligates with a subject that has the semantic preference to be human in authority. When used in the passive, it colligates specifically with /dâyráp/. It seems that the only genre-related variation that exists is in the nature of the active subject of /ʔànúyâat hây/, as discussed above.

5.5.14 /sǔɔn/ ‘teach’

/sǔɔn/ has two meanings: ‘teach’ and ‘practice’. The only meaning that appears across all four genres is ‘teach’.

When /sǔɔn/ means ‘teach’, it is a ditransitive verb. That is, it takes a direct object and an indirect object, and participates in various structures typical of ditransitive verbs. Most of the direct object nouns are abstract concepts. The indirect object has a semantic preference to be human. It is usually preceded by /hây/, /hây kàp/, or /hây kèɛ/. Here, /hây/ is a dative marker. /kàp/ and /kèɛ/ are prepositions meaning ‘with’. These markers are not specific to /sǔɔn/; they are general markers that often precede an indirect object.

Example 5.25 (*newspaper*)

thâa	phôw	pen	phôw	phǒm	thammay	mây	yɔɔm
if	father	COP	father	1SG	why	NEG	resolve
sǔɔn	wichaa		ʔaakhom		hây	phǒm	
teach	knowledge		magic		DAT	1SG	

‘If you are my father, why don’t you resolve to teach me your magic knowledge.’

Even though /sǔɔn/ is a ditransitive verb, a direct and indirect object are not always present at the same time. In fact, instances where both are present are much less frequent than those where only the direct object or the indirect object is present.

Example 5.26 (*fiction*)

nũu	pen	câawkhǔɔŋ	phaasǎa	khon	cà	sǔɔn	khǎw
2SG	COP	owner	language	should	CM	teach	3SG
dây	dii	khàa	chǎn				
POT	good	than	1SG				

‘You are a native speaker. You should be able to teach him better than me.’

In contrast to Example 5.25, in Example 5.26, there is no dative marker before the indirect object.

There are also 154 instances across the four genres where both direct and indirect objects are omitted, and must be inferred from context.

/sǒn/ may alternatively be followed by a clause preceded by the causative marker /hây/, creating a causative serialisation (see section 1.3.3.2). /sǒn hây/ occurs across all four genres.

Example 5.27 (*non-academic*)

thâan	yaŋ	sǒn	hây	pàtibàt
3SG	also	teach	CAUS	practice
phromwihǎan				4
principles.virtuous.existence				4

‘He also taught us to practice the four principles virtuous existence.’

/sǒn hây/ is arguably the core of an extended unit of meaning that has the pragmatic function of expressing that someone is teaching someone else to do something. This meaning is expressed across the unit, rather than by any individual word of the unit. This extended unit of meaning can be laid out as follows:

[person who teaches] /sǒn hây/ [person who is taught] [verb] ([object/adverb])

In conclusion, /sǒn/ is arguably most frequently used as a single-word unit of meaning, expressing the meaning of ‘teach’. With this use, /sǒn/ requires the colligations and semantic preferences typical of the word class it belongs to. As a ditransitive verb, it requires a subject, a direct object as well as an indirect object, the first and the third of which in turn have a semantic preference to be human. That said, the sequence /sǒn hây/ is arguably the core of an extended unit of meaning with a semantic prosody. There does not appear to be any variation in the use of /sǒn/ across genres.

5.5.15 /càtkaan/ ‘manage, deal with’

/càtkaan/ has a flexible meaning. On its own, it means ‘manage’, or ‘deal with’. But it also has other meanings, depending on context. The meaning of /càtkaan/ as ‘manage’ and ‘deal with’ is prevalent across all four genres.

When /càtkaan/ means ‘manage’ or ‘deal with’, it takes a noun object, often a direct object immediately after /càtkaan/, but sometimes a prepositional phrase object with /kàp/ ‘with’. The nouns are mostly abstract, and most of them are semantically neutral. One frequent neutral object noun is /râaŋ/ ‘matter’. There are 10 positive and 41 negative instances of object nouns. Two frequently co-occurring positive and negative object nouns are /sáppáyaakoon/ ‘resources’, and /panhãa/ ‘problem’, respectively. These co-occurrences are arguably collocations. The meaning each sequence conveys is no more than the combination of the meanings of /càtkaan/ and the object noun that follows it.

When /càtkaan/ takes a noun as object, with or without /kàp/, the meaning the combination expresses is not always ‘manage’ or ‘deal with’. When /càtkaan/ takes a human object, the meaning of the combination varies but is almost always negative. There are only three instances of neutral meaning, and no instances of positive meaning. The combination of /càtkaan/ plus a human object can mean ‘get rid of someone’, ‘hurt someone’, or ‘take revenge on someone’, depending on context. This type of combination occurs 19 times in *fiction* and 12 times in *non-academic*. There are only four instances in *newspaper* and none in *academic*.

Example 5.28 (*fiction*)

càtkaan	dèk	nân	mây	tôŋ	thǎŋ	taay	khêε
manage	kid	DEM	NEG	must	until	die	just
phɔɔ	hây	man	khâw	pay	yùu	roŋpháyaabaan	
enough	CAUS	3SG	enter	DIR	stay	hospital	
sàk	sîi	duan					
about	four	month					

‘Make that kid hurt, but don’t kill him. Just make him have to stay in the hospital for about four months.’

In *fiction*, there are 13 instances where /càtkaan/ is followed by a food item, which may be preceded by /kàp/. In such contexts, /càtkaan/ means ‘eat/finish/prepare (the food)’.

/càtkaan/ may be immediately followed by /ʔeeŋ/. /ʔeeŋ/ is an adverb, meaning ‘by oneself’. There are 22 instances of /càtkaan ʔeeŋ/ in *fiction* and 1 in *newspaper*. Interestingly, /càtkaan ʔeeŋ/ almost always appears at the end of the sentence (a colligation). In all instances, the subject of /càtkaan ʔeeŋ/ is the speaker, and the sequence is always used in a context where the speaker proposes or insists on managing or dealing with something for the interlocutor. Usually, the speaker asks the interlocutor to do or not to do something before or after offering to manage the matter him/herself.

Example 5.29 (*fiction*)

mây	tôŋ	khà	kô	bòok	léew	ŋay
NEG	must	SLP	LP	tell	ASP	PP
cíip	càtkaan		ʔeeŋ			
Jib	manage		by.oneself			

‘You don’t have to help me. I told you. I will manage it myself.’

/càtkaan ʔeeŋ/ is therefore arguably the core of an extended unit of meaning that has a grammatical preference to appear at the end of the sentence and the pragmatic function of rejecting (possibly in advance) an offer of help. This pragmatic meaning is

spread across the sequence and is often implied in context. This extended unit of meaning can be laid out as follows:

([expression of request]) | [speaker self-reference] /càtkaan ʔeeŋ/ | ([expression of request])

/càtkaan/ may also appear in a serial verb construction. The meaning the sequence of /càtkaan/ plus a verb conveys is ‘manage to do something’.

Example 5.30 (*fiction*)

phôw	bòn	nít	nít	léew	càtkaan	banthúk
father	complain	little	little	LINK	manage	record
khôomuun	loŋ	pay	yàaŋ	rûatrew		
information	ASP	ASP	AZP	quick		

‘Father complained a little bit, and then managed to quickly record the information.’

/càtkaan/ may also be followed by the causative marker /hây/, creating a causative serialisation (see section 1.3.3.2). Mostly, the resulting event is not present after /hây/, but must be inferred from context.

Example 5.31 (*fiction*)

raw	cà	chûay	càtkaan	hây		khun mây	tôŋ
1PL	CM	help	manage	CAUS		2SG NEG	must

penhùaŋ
worry
‘We will help arrange that (for you). You don’t have to worry.’

/càtkaan hây/ is arguably the core of an extended unit of meaning that has the pragmatic function of expressing an arrangement for something to happen. This meaning is spread pragmatically across the sequence, rather than evident from any individual word of the sequence. Interestingly, when the subject of /càtkaan hây/ is first person, the pragmatic function the sequence expresses is not only an arrangement, but is also likely to be an offer to make that arrangement, as seen in Example 5.31. Frequently, what is

arranged is implicit, and must be inferred from context. That is, in this structure /hây/ has no clausal complement. This extended unit of meaning can be laid out as follows:

[person] /càtkaan hây/ |

There are as many as 22 instances of /càtkaan hây/ in *fiction*, but only 4 in *newspaper*, and 3 in each of *academic* and *non-academic*.

In sum, we see that /càtkaan/ is most frequently used as a single-word unit of meaning, expressing no more than the meaning of ‘manage’ or ‘deal with’. When used this way, /càtkaan/ requires colligations and semantic preferences typical of this type of verb. It colligates with a subject, which has a semantic preference to be human, as well as an object, which tends to be abstract. There are also instances, especially in *fiction* (19) and *non-academic* (12), where the object of /càtkaan/ is human. That said, /càtkaan/ also forms part of the core of two different extended units of meaning, as we have seen with /càtkaan ʔeeŋ/ and /càtkaan hây/, each of which has a clear semantic prosody. We have also seen some variation across genres. The sequences of /càtkaan/ plus human object, /càtkaan ʔeeŋ/ and /càtkaan hây/ occur frequently in *fiction*, but not in the other three genres. In fact, /càtkaan ʔeeŋ/ does not occur at all in *academic* and *non-academic*.

5.5.16 /hăay/ ‘recover, disappear’

/hăay/ has two meanings: ‘recover (from illness)’ and ‘disappear’. It can also appear in a serial verb construction. When used in a serial verb structure, /hăay/ arguably forms part of an extended unit of meaning. These three uses of /hăay/ are prevalent across all four genres.

When /hǎay/ means ‘recover (from illness)’, it is frequently followed by /dii/ ‘well’, /rɛw/ ‘quickly’, /cháa/ ‘slow’, or /pùay/ ‘sick’. /hǎay dii/, /hǎay rɛw/, /hǎay cháa/, and /hǎay pùay/ mean ‘fully recover’, ‘recover quickly’, ‘recover slowly’ and ‘recover (from illness)’, respectively. These sequences are arguably collocations. The meaning they convey is no more than the combination of the meanings of each element of the sequence. /hǎay/ may also be followed by a prepositional phrase beginning with /càak/ ‘from’ (a colligation), which is in turn followed by /rôok/ or /rôok phay khây cèp/, both of which mean ‘disease’, or the name of a disease. Even though this sequence is rather fixed, I would argue that it is not an extended unit of meaning, because it has a fully compositional meaning of recovering from a (particular) disease.

Example 5.32 (*academic*)

khonkhây	kô	cà	phlɔɔy	mii	sùkhàphâapcìt	thii
patient	LP	CM	accordingly	have	mental.health	SBR
dii	thamhây	hǎay	càak	rôok	phay khây	cèp
good	cause	recover	from	disease	danger	fever
					hurt	

‘Patients will accordingly have good mental health, which makes them recover from illness.’

There are three instances of this sequence in *academic*, four in *newspaper*, two in *non-academic*, and none in *fiction*.

With the meaning of ‘recover (from illness)’, the subject of /hǎay/ can be human, or an expression of physical damage or disease, such as /phlǎɛ/ ‘wound’ and /ʔaakaan/ ‘symptom’.

When /hǎay/ means ‘disappear’, it is most frequently followed by /pay/. /pay/ literally means ‘go’, but is also an aspect marker indicating completion (Iwasaki and Ingkaphirom 2005: 163). /hǎay pay/ means ‘disappear completely’. The subject of /hǎay pay/ can be abstract or concrete, including human. With this meaning, /hǎay/ may be used

in the sense of ‘something being lost’, although this latter use is less frequent than the sense of ‘disappearing’.

Example 5.33 (*academic*)

dontrii	thay	nay	pàtcùban	thêep	cà	hăay
music	Thai	in	present	almost	CM	disappear
pay	càak	săᅅkhom	thay			
ASP	from	society	Thai			

‘Thai music has almost disappeared from the Thai society these days.’

/hăay/ can alternatively be used in a serial verb structure. When used this way, interestingly, almost all the verbs following /hăay/ are semantically negative; there is a semantic preference for physical states or emotions. Negative verbs that occur repeatedly include /khriat/ ‘stressed’, /nùay/ ‘tired’, /bùa/ ‘bored’, /sõᅅsăy/ ‘curious’, and /ᅅoᅅ/ ‘confused’. The combination of /hăay/ plus a negative verb yields a combination with the positive meaning of ‘no longer be stressed/tired/bored/ etc.’ There is one neutral verb in the data: /yàak/ ‘have an appetite for’. However, the sequence /hăay yàak/ seems to have a positive meaning. This use of /hăay/ plus a verb in a serial verb construction occurs across all four genres, most frequently in *fiction* (31 instances). In this use, the subject of /hăay/ is always human.

Example 5.34 (*fiction*)

phùakthəə	man	ᅅôo	mûarày	cà	hăay	ᅅôo
2PL	2PL	foolish	when	CM	disappear	foolish

‘You are foolish. When will you stop being foolish?’

Here, /hăay/ arguably is the core of an extended unit of meaning. It has a colligation for a following verb, which in turn has a semantic preference to express an attribute (a negative physical state or emotion). Overall, the unit has the pragmatic function of expressing unpleasant emotions or physical states coming to an end. This

meaning cannot be derived from any individual word in the unit, but is rather expressed pragmatically across the sequence. The unit can be laid out as follows:

[person] /hăay/ [verb expressing (negative) physical state or emotion]

We have seen that /hăay/ is most frequently used as a single-word unit of meaning, whether it means ‘recover’ or ‘disappear’. When so used, it requires colligations and semantic preferences typical of the relevant type of verb. When it means ‘recover (from illness)’, it requires a subject which in turn has the semantic preference to be human. When it means ‘disappear’, it also requires a subject. However, in this context, the subject may also be non-human, or abstract. That said, /hăay/ arguably is also the core of an extended unit of meaning when it is used in a serial verb structure. There is not any clear variation across genres.

5.5.17 /baŋkháp/ ‘force, execute, steer’

/baŋkháp/ on its own means ‘force’, ‘execute (a judgement)’, or ‘steer (a vehicle)’. The meaning of ‘force’ is prevalent across all four genres. Another meaning, ‘execute (a judgement)’, appears only in *academic* and *newspaper*. There are eight instances of /baŋkháp/ meaning ‘steer’ in *fiction*, and two in each of *newspaper*, and *non-academic*. Other than these two meanings, /baŋkháp/ also forms part of some fixed phraseologies.

When /baŋkháp/ means ‘force’, it takes a noun as direct object, which can be either abstract or concrete, but is most frequently human. /baŋkháp/ may alternatively be followed by a clause preceded by /hây/, creating a causative serialisation (see section 1.3.3.2). /baŋkháp hây/ appears in both active and passive structures, although it is used

more frequently in an active structure. This applies across all four genres. There are also six instances of /baŋkháp mây hây/, where /mây/ is a negator. However, /baŋkháp mây hây/ occurs only in the active structure.

Example 5.35 (*fiction*)

chăn	baŋkháp	hây	khăw	còt	rót
1SG	force	CAUS	3SG	park	car

‘I forced him to stop the car.’

When active, the subject of /baŋkháp (mây) hây/ (the causer) is almost always human. There are 51 non-human subjects, e.g. /kòtmăay/ ‘law’, and /sàthănákaan/ ‘situation’. The subject of the clause following /hây/ (the causee) is also almost always human, with 25 exceptions where it is non-human, such as /mũu/ ‘hand’, and /sĩaŋ/ ‘voice’. Frequently, the causee appears after /hây/, as seen in Example 5.35. However, there are also 30 instances where the causee appears immediately after /baŋkháp/ and before /(mây) hây/, as shown in Example 5.36.

Example 5.36 (*academic*)

bùkkhon	thĩ	mây	dây	pen	tháhăan	nán		
person	SBR	NEG	POT	COP	soldier	PP		
cà	khum	tháhăan	wáy	mây	dây	rũu	mây	mii
CM	control	soldier	ASP	NEG	POT	or	NEG	have
ʔamnâat	phoo	thĩ	cà	baŋkháp	tháhăan			
power	enough	SBR	CM	force	soldier			
hây	chũafan	dây						
CAUS	obey	POT						

‘Those who are not a soldier can’t control soldiers, or they don’t have enough power to force soldiers to obey them.’

When used in a passive structure, /baŋkháp hây/ colligates specifically with the passive auxiliaries /thùuk/ and /doon/, which come before /baŋkháp hây/. Unlike the other passive auxiliary /dâyráp/, /thùuk/ and /doon/ usually indicate adverse situations (Iwasaki

and Ingkaphirom 2005: 313). There may be a reference to the causer (a person) between the passive auxiliary and /baŋkháp hây/.

Example 5.37 (*newspaper*)

thay	kamlan	thùuk	baŋkháp	hây
Thailand	ASP	PASS	force	CAUS
lót	kaan	sòŋʔòok		
reduce	NMLZ	export		

‘Thailand is being forced to reduce exportation.’

As in the active, the subject of /baŋkháp hây/ (the causee) is almost always human.

/baŋkháp (mây) hây/, whether active or passive, is arguably the core of an extended unit of meaning that has the pragmatic function of expressing that someone or something makes someone (else) do something or make something happen. This meaning, that is, the expression of causation via a periphrastic causation, is spread pragmatically across the unit. Thus, we see that the unit has its own specific semantic preference and colligation. In both active and passive, the causer and the causee have a semantic preference to be human. When passive, as previously discussed, /baŋkháp hây/ specifically colligates with /thùuk/ and /doon/. The unit of meaning whose core is /baŋkháp (mây) hây/ can be laid out as follows:

Active structure

[person/thing that forces] /**baŋkháp (mây) hây**/ [person/thing that is forced] [verb] ([object/
adverb])

[person/thing that forces] /**baŋkháp**/ [person/thing that is forced] /**(mây) hây**/ [verb] ([object/
adverb])

Passive structure

[person/thing that is forced] /thùuk/ [person/thing that forces] /**baŋkháp hây**/ [verb] ([object/
/doon/ adverb])

This unit of meaning is found across all four genres. It is the most frequent use of /baŋkháp/ in *fiction* (88 instances) and *non-academic* (63 instances). There are 40 instances of this use in *newspaper* and 24 in *academic*.

When /baŋkháp/ means ‘execute (a judgement)’, it also takes a noun as direct object. One frequent object noun is /khádii/ ‘(legal) case’. /baŋkháp khádii/ means ‘execute a judgement’; it is arguably a collocation. The meaning the sequence conveys is no more than the combination of the meanings of the elements of the parts.

/baŋkháp/ also forms part of some fixed phraseologies, including /mii phõn baŋkháp/, /mii phõn baŋkháp cháy/, /mii phõn cháy baŋkháp/, /cháy baŋkháp/, and /baŋkháp cháy/.

/mii phõn baŋkháp/, literally ‘have result force’, /mii phõn baŋkháp cháy/, literally ‘have result force use’, and /mii phõn cháy baŋkháp/, literally ‘have result use force’, all mean ‘become effective’. These phraseological units function as intransitive verbs. Their subject has a semantic preference to refer to rules, especially in legal contexts, with nouns such as /kòtmăay/ ‘law’, /phráráatchábanyàt/ ‘act’, or /mâattraa/ ‘section’.

Example 5.38 (*newspaper*)

kòtmăay	chàbàp níi	mii	phõn	baŋkháp	cháy	yàaŋ
law	CLS	DEM	have	result force	use	AZP
pen	thaaŋkaan	maa	tâŋtèè	wanthîi	11	meesăayon
COP	official	ASP	since	day	11	April
2011						
2011						

‘This law came into force since the 11th of April 2011.’

/cháy baŋkháp/, literally ‘use force’, and /baŋkháp cháy/, literally ‘force use’, mean ‘enforce’. What is enforced also has a semantic preference for regulations, such as

/kòtmăay/ ‘law’ and /mâattraa/ ‘section’. This noun may come before or after /cháy bạ̀kháp/ and /bạ̀kháp cháy/.

Example 5.39 (*academic*)

mii	khwaampháyaayaam	thi	cà	traa	kòtmăay
have	attempt		COMP CM	enact	law
kiawkàp	khwaam	plòtphay	nay	kaan	thamṇaan
about	NMLZ	safe	in	NMLZ	work
dooychàpó	khên	maa	cháy	bạ̀kháp	
particularly	ASP	ASP	use	force	

‘There was an attempt to enact a law particularly about safety in work and enforce it.’

These fixed phraseological units are arguably variant options for the core of two separate extended units of meaning, which can be laid out as follows:

[regulation] /**mii phôn bạ̀kháp/**
 /**mii phôn bạ̀kháp cháy/**
 /**mii phôn cháy bạ̀kháp/**

[regulation] /**cháy bạ̀kháp/** [regulation]
 /**bạ̀kháp cháy/**

Both extended units of meaning have a clear pragmatic function, of expressing the coming into effect of a regulation and the enforcement of a regulation, respectively. These pragmatic functions or semantic prosodies are pragmatically spread across the sequence, rather than solely attributable to any element of the unit. The use of these units seems to be specific to legal contexts. There are 54 instances of these extended units in *academic* and 67 in *newspaper*. There are 16 in *non-academic*, but none in *fiction*.

In sum, /bạ̀kháp/ is frequently used as a single-word unit of meaning with colligations and semantic preferences typical of this type of transitive verb. When it means ‘force’, it requires a human subject and an object, which can be either abstract or concrete, but mostly human. When it means ‘execute’, it also requires a human subject as

well as an object. That said, with the meaning of ‘force’, /baŋkháp/ also arguably forms part of an extended unit of meaning, as we have seen with /baŋkháp (mây) hây/. It also forms part of another two separate extended units of meaning, as previously discussed. These latter two extended units of meaning seem to be restricted to legal contexts and appear only in *academic*, *newspaper*, and *non-academic*, but not at all in *fiction*.

5.5.18 /duulɛɛ/ ‘take care of’

/duulɛɛ/ means ‘take care of’. It is transitive and takes a noun as direct object, often immediately after /duulɛɛ/. This applies across all four genres. The object noun can be either abstract or concrete, although the abstract nouns outnumber the concrete nouns. Many of the abstract nouns are marked by /kaan/ or /khwaam/, the grammatical particles creating an abstract noun. The object noun can also be human. Human object nouns occur across all four genres, but especially in *fiction*, where it is the most frequent type of object. There are also 57 instances where the object noun is not present, but must be inferred from context.

/duulɛɛ/ may be followed by /tuaʔeeŋ/ or /tonʔeeŋ/, both of which are reflexive pronouns. /duulɛɛ tuaʔeeŋ/ and /duulɛɛ tonʔeeŋ/ mean ‘take care of oneself’. There are eight instances of these colligations in each of *non-academic* and *newspaper*, and eleven in *fiction*, but only two instances in *academic*. /duulɛɛ tuaʔeeŋ/ and /duulɛɛ tonʔeeŋ/ may be followed by the potential auxiliary /dây/, literally ‘be able to’. /duulɛɛ tuaʔeeŋ dâŋ/ and /duulɛɛ tonʔeeŋ dâŋ/ mean ‘be able to take care of oneself’. Even though these sequences are somewhat fixed, they are better seen as colligation of /duulɛɛ/ with reflexivity, rather as extended units of meaning. The meaning the sequences express is no

more than the combination of the meanings of the individual items of the sequence. There are no instances of the phraseology with both reflexive and auxiliary in *academic*, one in *newspaper*, two in *non-academic*, but as many as seven in *fiction*.

/duuleɛ/ also occurs in a serial verb construction. It can come in either first or second position in the structure. Interestingly, the other verbs it combines with tend to have similar meanings to /duuleɛ/, such as /pòkpôŋ/ ‘protect’, /rápphitchôp/ ‘responsible’, and /ráksǎa/ ‘protect’. This seems to represent a type of semantic reduplication common in Thai (Iwasaki and Ingkaphirom 2005: 36). These sequences (with /duuleɛ/ either first or second) occur repeatedly and arguably form collocations.

Example 5.40 (*academic*)

khàndiawkan	sǎŋkhom	kô	mii	phaarákìt	thîi	cà	
meanwhile	society	LP	have	mission	COMP	CM	
tôŋ	duuleɛ	ráksǎa	sáppháyaak	ɔnthammá	châat		
must	take.care	protect	natural	resources			

‘Meanwhile, the society also has a mission to take care of natural resources.’

/duuleɛ/ may be followed by a clause preceded by the causative marker /hây/, creating a causative serialisation (see section 1.3.3.2).

Example 5.41 (*non-academic*)

ʔiik	sip	pii	khâaŋnâa	phôm	khon	cà	pen
another	ten	year	front	1SG	will	CM	COP
phô	khôŋ	lúuk	tua	nóoy	thîi	cà	tôŋ
father	of	child	body	small	SBR	CM	must
duuleɛ	hây	khǎw	yâaŋ	kâaw	sùu	wayrûn	
take.care	CAUS	3SG	step	step	to	adolescence	
dúay	khwaam	rámátráwaŋ					
with	NMLZ	careful					

‘In ten years’ time, I will be a father of a small child who has to ensure that my child will carefully step into adolescence.’

The sequence /duuleɛ hây/ is arguably the core of an extended unit of meaning that has the pragmatic function of expressing an assurance that something that should or should not happen does indeed happen or not happen. This meaning is not evident from any individual element of the unit, but rather spread pragmatically across the whole sequence, which is thus the unit's semantic prosody. The negator /mây/ may occur between /duuleɛ/ and /hây/. The unit can be laid out as follows:

[person] /**duuleɛ (mây) hây**/ [another person/thing] [verb] ([object/adverb])

There are not many instances of this unit across all four genres: five in each of *academic*, *newspaper*, and *non-academic*, and one in *fiction*.

In sum, /duuleɛ/ is arguably most frequently used as a single-word unit of meaning expressing the meaning of 'take care of (something or someone)'. With this use, it requires colligation and semantic preference typical of this type of verb. But it also has its own specific collocations – the semantic reduplications. As a transitive verb, it requires a subject which has the semantic preference to be human as well as an object, which need not be human. It also colligates with the reflexive pronouns and the potential auxiliary in one common phraseology. That said, there are also instances where /duuleɛ/ forms part of the core of the extended unit of meaning with a clear pragmatic function, as we have seen with /duuleɛ (mây) hây/. There does not seem to be any obvious variation across genres.

5.5.19 /còp/ ‘graduate, end/finish’

/còp/ means ‘graduate’ or ‘end/finish’. It can also function as an aspect marker indicating completion (Iwasaki and Ingkaphirom 2005: 160). These three uses of /còp/ are prevalent across all four genres.

When /còp/ means ‘graduate’, it may be followed by /kaansùksǎa/ ‘education’, which may in turn be followed by the level of education. /còp/ may alternatively be immediately followed by the level of education without /kaansùksǎa/ in between. Often, however, /còp/ is used on its own. With this meaning, the subject of /còp/ is always human.

When it means ‘end/finish’, /còp/ can be used either intransitively or transitively. That is, it can mean ‘something coming to an end’ or ‘something ending something else’. The intransitive use is much more frequent than the transitive use across all four genres. When /còp/ is used intransitively, it is often followed by the aspect marker /loŋ/, literally ‘descend’. /còp loŋ/ is prevalent across all four genres. There are also instances where /còp/ is followed by /sîn/, meaning ‘end’, or /kan/, a reciprocal marker. /còp sîn/ occurs 14 times across all four genres, but /còp kan/ occurs only 5 times in *academic*, *fiction*, and *non-academic*. /còp loŋ/, /còp sîn/, and /còp kan/ all mean ‘come to an end.’ Nevertheless, whereas /còp loŋ/ and /còp sîn/ can occur in any context, /còp kan/ seems to be restricted to negative contexts. In fact, as a native speaker of Thai, I have the impression that /còp kan/ is the core of an extended unit of meaning that has the pragmatic function of expressing an unpleasant, unavoidable ending. However, this pragmatic function is not clearly expressed in all instances of /còp kan/. Of the five instances of /còp kan/ across all four genres, there are three instances where this

pragmatic meaning is explicitly expressed. Therefore, I cannot state confidently that /còp kan/ is the core of an extended unit of meaning that has the pragmatic function of expressing an unpleasant, unavoidable ending. To argue for this, more examples would be needed.

Example 5.42 (*fiction*)

thâa	phlâat	khraaw	nîi	kô	penʔan	còp	kan
if	fail	time	DEM	LP	mean	end	REC

‘If (I) fail this time, that means it has to come to an end.’

Example 5.43 (*newspaper*)

duu	măanwâa	rûaŋ	nâa	cà	còp	loŋ	dûay
look	as.if	story	should	CM	end	ASP	with
dii	têe	hêtkaan	kô	klàp	mây	pen	chênnán
good	but	event	LP	return	NEG	COP	like.that

‘It looked as if the story would end well, but it didn’t turn out like that.’

Example 5.44 (*academic*)

mée	kaan	bèeŋyêek	ráwàaŋ	sáay	kàp	khwăa
although	NMLZ	separate	between	left	with	right
cà	mòt	pay	têe	kaan	kh-rûnkhít	thŋŋ
CM	terminate	ASP	but	NMLZ	muse	to
panhăa	khwaam	thâwthiam	yàaŋ	ciŋcaŋ	kô	mây
problem	NMLZ	equal	AZP	serious	LP	NEG
campen	tôŋ	còp	sîn			
necessary	must	end	end			

‘Although the separation between the Left and the Right no longer exists, musing on problems about equality seriously doesn’t have to end.’

However, when /còp sîn/ is preceded by the negator /mây/, as in /mây còp sîn/, /mây mii wan còp sîn/, literally ‘not have day end end’, or /mây rúucàk còp sîn/, literally ‘not know end end’, it seems to be restricted to negative contexts. Out of 10 instances, there are as many as 9 instances where each of these sequences is used in negative contexts, and only one in a neutral context. /mây còp sîn/, /mây mii wan còp sîn/, and /mây rúucàk còp sîn/ are thus arguably variant options for the core of an extended unit of

meaning that has the semantic prosody of expressing something unpleasant continuing indefinitely, which can be laid out as follows:

[something unpleasant] /mây còp sîn/
 /mây mii wan còp sîn/
 /mây rúucàk còp sîn/

Example 5.45 (*newspaper*)

nay	ʔànaakhót	panhăa	khwaamkhàtyéεη	cà	mây
in	future	problem	conflict	CM	NEG
còp	sîn				
end	end				

‘In the future, problems of conflict will never end.’

When used transitively, the subject of /còp/ tends to be human. One frequent object of /còp/ is /chiiwít/. /chiiwít/ means ‘life’. /còp chiiwít/, literally ‘end life’, means ‘die’. There are four instances of /còp chiiwít/ in *fiction*, and three in each of *newspaper* and *non-academic*, but none in *academic*. /còp chiiwít/ seems to me to be the core of an extended unit of meaning that has the pragmatic function of expressing that death comes too soon or is unexpected. However, this pragmatic meaning is not reflected in all instances of /còp chiiwít/. It is expressed in only five examples, where the pragmatic meaning of unexpected or premature death is implied by the context.

Example 5.46 (*fiction*)

naay	cà	maa	còp	chiiwít	loη	dúay
2SG	CM	ASP	end	life	ASP	with
ʔùbàttihèet	mæw	mæw	bèεp	níi	mây	dây
accident	cat	cat	AZP	DEM	NEG	POT

‘You must not die from this tiny silly accident.’

As not all the instances of /còp chiiwít/ have a clear pragmatic function distinct from the meaning of the sequence, I cannot state with confidence that /còp chiiwít/ is the

core of an extended unit of meaning, although my intuition as a native speaker tells me so. To argue for this, more examples would be needed.

When /còp/ is used as an aspect marker, it is often immediately preceded by a verb. There are also 34 instances where /con/ ‘until’ or the causative marker /hây/ occurs between the verb and /còp/. However, these are much less frequent than those where the verb and /còp/ are adjacent (157 instances). Frequently co-occurring verbs include /phaŋ/ ‘listen’, /phûut/ ‘speak’, /ʔaan/ ‘read’, and /khĩan/ ‘write’. These verbs arguably form colligations with /còp/. The meaning these sequences express is ‘finish doing something’, such as ‘finish listening’ for /phaŋ còp/, and ‘finish speaking’ for /phûut còp/. With this use, the subject of /còp/ is always human.

In sum, whether it means ‘graduate’ or ‘end/finish’, /còp/ is mostly used as a single-word unit of meaning and requires the colligations and semantic preferences typical of this type of verb. When it means ‘graduate’, /còp/ requires a subject which in turn has the semantic preference to be human. When it means ‘end/finish’, it is used either transitively or intransitively, with the appropriate subject/object colligations. That said, there are also instances where /còp/ arguably forms part of the core of an extended unit of meaning with a clear semantic prosody, as we have seen with /mây còp sîn/, /mây mii wan còp sîn/, and /mây rúucàk còp sîn/. /còp kan/ and /còp chiiwít/ seem to me to be the cores of two further extended units of meaning, but to argue for this confidently, more examples would be needed. There is not any obvious variation across genres.

5.6 Discussion

In this section, I will first pick up on some issues that have repeatedly come up throughout the analysis. Then, I will discuss variation in semantic prosodies across genres emerging from the data, which is the main aim of the analysis.

5.6.1 Single-word units of meaning

We have seen throughout the analysis that each of the verbs under study, except /hǎn/ ‘turn’, is most frequently used as a single-word unit of meaning. However, even used this way, each verb still requires colligations and semantic preferences that are typical of the broad class of verb it belongs to. For example, /pràkàat/, as a reporting verb meaning ‘announce’ or ‘declare’, requires a subject that has a semantic preference to be human as well as an object. These general patterns do not appear to create an extended unit of meaning in Sinclair’s sense, but they do constitute specific patterns of these verbs.

These general patterns identified when the verb is used as a single-word unit of meaning would seem to correspond to the patterns that Hunston and Francis (2000) discuss in terms of Pattern Grammar, their systematisation of the syntactic structure that appears around particular English words. Hunston and Francis (2000: 37) define the patterns of a word as “all the words and structures which are regularly associated with the word and which contribute to its meaning’. Even though patterns really belong to a particular word, Hunston and Francis (2000: 43) argue, the patterns are described as general, rather than specific, syntactic patterns. In other words, they are patterns that co-occur with a whole set of words, not descriptively specific to any individual word.

This view of patterns is congruent with what we have discovered in the analysis. We have seen, for example, a colligational pattern of a verb followed by a noun, which can be laid out as V n (verb plus noun) using the notation of Pattern Grammar. I observed this pattern to be associated with all the verbs that are transitive, such as /càp/, /khûapkhum/, and /níyom/, as this is the syntax typical of a direct object (although in Thai, in a minority of cases, the direct object noun can precede the verb, as I have mentioned). The colligational pattern of a verb followed by a verb, which can be laid out in Pattern Grammar style as V v, is associated with all the verbs that frequently appear in a serial verb construction, such as /pràkàat/, /yɔɔm/, and /yuaan/.

Similarly, Pattern Grammar holds that a word can also have several patterns (Hunston and Francis 2000: 83). We have seen throughout the analysis that each verb under study has more than one pattern. /duuleɛ/, for example, has both of the patterns V n and V v, that is, it can be followed by a noun as direct object or by a verb in a serial verb structure. Hunston and Francis (2000: 83) further argue that in cases where a word has more than one meaning, the pattern in which the word occurs determines what meaning is being used. This point is again congruent with what we have discovered in the analysis. For example, /waan/ has four different meanings, two of which are ‘put’ and ‘be placed’. When /waan/ means ‘put’, it occurs in the pattern V n, that is, it is followed by a noun as direct object, and has a semantic preference for a human subject. When it means ‘be placed’, by contrast, it appears before the aspect marker /yuu/, which can be laid out as V /yuu/, and has a semantic preference for its subject to be a concrete noun when used literally and an abstract noun when used metaphorically. This underlines not only Hunston and Francis’ point that the different meanings of the word tend to be associated

with different syntactic structures, but also the point that the semantic preferences of a word that has more than one meaning may vary according to the meaning it conveys.

So far I have argued that the patterns that I identified when the verb is used as a single-word unit of meaning correspond to the patterns Hunston and Francis discuss in terms of Pattern Grammar. In fact, I would go further and argue that the very idea of a single-word unit of meaning is compatible with the Pattern Grammar view. This is because, as I noted many times in the analysis, even when a verb is used as a single-word unit of meaning, it requires the colligations and semantic preferences typical of the class of verb it belongs to. These class-linked colligations/semantic preferences *are* what Hunston and Francis describe as patterns. The verb /càp/ ‘touch’, for example, when used in its literal sense, has a requirement for an object (a colligation) that has to be concrete (a semantic preference). This creates a general pattern for /càp/, which can be presented as V n in the notation of Pattern Grammar. When I refer to a verb as being a single-word unit of meaning, the unit of meaning in question is the verb plus one of its associated general patterns (such as V n). It is a *single-word* unit even so because the words that *realise* the other elements of the pattern are not themselves part of that unit of meaning. Instead, they are another separate (single-word) unit of meaning linked to the original unit of meaning by the colligations and semantic preferences required by the verb. For example, consider this example of /càp/ used as a single-word unit of meaning in the pattern V n: /ʔaakhirá càp lày chǎn/ ‘Akira touched my shoulder’. Here, /lày/ ‘shoulder’ is a separate single-word unit of meaning, but it fulfils the noun slot in the pattern V n, which is part of the unit of meaning based on /càp/. In other words, /lày/ fulfils the general colligation and semantic preference required by the transitive verb /càp/ for its

object. Therefore, the overall meaning of the sequence of the verb plus the words that fulfil its pattern is the sum of the various single-word units of meaning – that is, the verb and its pattern plus the other words nearby that realise the required colligations and semantic preferences. As Hunston and Francis (2000: 86) note, a pattern can be viewed as a framework of slots into which speakers insert particular words, albeit this is not their preferred view. But we see that this view in fact fits well with the Thai data, where the verb plus its pattern can be viewed as a single-word unit of meaning, and the overall meaning of the sequence (verb plus pattern, and the other words fulfilling the pattern) is simply compositional.

Thus, I have argued that when used as a single-word unit of meaning, even with an identifiable grammar pattern, the verb expresses no more than the meaning of the word considered in isolation. The pragmatic function of the verb is to convey the meaning the verb has. Nor does the combination of the verb with the elements that fulfil its patterns (colligation, semantic preference) go beyond the combination of the meanings of that set of items. Therefore, overall these patterns do not create an *extended* unit of meaning in the Sinclairian sense, for they do not exhibit a pragmatic function, i.e. a semantic prosody in Sinclair's sense, and only require general colligations and semantic preferences typical of the verb type they belong to. This justifies the label "single-word" unit of meaning rather than "extended" unit of meaning for these usages.

As well as occurring within patterns in the sense of Pattern Grammar, a verb may participate in collocations when it is being used as single-word unit of meaning. The meaning that these collocations express is not clearly distinct from a combination of the

meanings of the verb and the collocate. Moreover, the combinations do not exhibit a clear pragmatic function and are not restricted to a particular context.

That said, I would argue that there do exist instances where the verb of such a collocational combination is delexicalised. Delexicalisation, as Philip (2011: 83) defines it, “refers to the gradual weakening of meaning that occurs when words are habitually used in combination.” When the verb participates in a collocation within which it is delexicalised, the meaning is mainly derived from the other element of the combination, rather than the verb. We have arguably seen this phenomenon with the verbs /càp/ ‘touch’, /waanj/ ‘put’, and /yók/ ‘lift’. When these verbs are used in their metaphorical sense alongside a direct object, the meaning of the combination is mainly contingent upon the direct object. For example, I earlier argued that when /càp/ means ‘touch’ and is used metaphorically, the meaning of the combination where /càp/ occurs depends primarily on the object noun. So, whereas /càp pràden/ means ‘grasp a point’, /càp khwaamrúusùk/ means ‘sense a feeling’.

5.6.2 Extended units of meaning

Some of the verbs studied do, however, also form part of one or more extended units of meaning. Out of 19 verbs, 14 arguably form (part of) the core of one or more extended units of meaning. These extended units of meaning require specific, rather than general, colligations and semantic preferences. They also have a clear pragmatic function or semantic prosody that is spread across the overall unit and is not equal to a combination of the meanings of the parts (see section 3.4.2). For example, the extended unit that has /càp dâj/ as its core has a specific colligation for the complementizer /wâa/

and a semantic preference for (bad) secrets. Overall, the unit has the pragmatic function of expressing that a (bad) secret is revealed. This overall meaning is spread across the whole sequence, rather than coming from any individual element. Table 5.2 lists all the extended units of meaning that I identified in the data and their semantic prosodies.

Verb	Extended units of meaning	Semantic prosodies
/pràkàat/	[person with authority] / pràkàat hây / [person/thing] [verb] ([object/adverb])	Announcement in which a person, particularly one with authority, asserts that he/she will make something happen
/càp/	[person or organisation] / càp muu / (/kàp/) [person or organisation] [verb] ([object/adverb]) [person] / càp / [another person] [verb] ([object/adverb]) [person] / càp dây / (/wâa/) ([secret/bad thing that has been hidden]) [person] /thùuk/ [another person] / càp dây / (/wâa/) ([secret/bad thing that has been hidden]) /doon/	Cooperation in doing something between two (groups of) people or organisations Someone making someone else do something (Bad) secret being revealed (active form) (Bad) secret being revealed (passive form)
/hân/	[person] / hân / /klàp maa/ [verb] ([object/adverb]) /klàp pay/ /maa/ /pay/	Someone starting to do something
/yɔɔm/	[person] / yɔɔm hây / [another person/thing] [verb] ([object/adverb])	Someone allowing someone else to do something or something to happen
/yók/	[some phenomenon] / yók tuayàaη chên / [example of phenomenon]	Linking a statement of some topic or phenomenon to an example of that topic or phenomenon
/khûapkhum/	[person who confines] / khûapkhum tua / ([person who is confined]) [person who is confined] /thùuk/ [person who confines] / khûapkhum tua / /doon/	Someone being confined (active form) Someone being confined (passive form)
/rɔɔ/	/ rɔɔ kaan kamnòt thòot / /rɔɔ kaan loηthòot/ /rɔɔ loη ʔaayaa/	Suspending determination of punishment in legal contexts Suspending a sentence in legal contexts Suspending a sentence in legal contexts
/ʔànúyâat/	[person who allows] / ʔànúyâat hây / [person who is allowed] [verb] ([object/adverb]) [person] /dâyrap/ / ʔànúyâat hây / [verb] ([object/adverb])	Permission being granted (active form) Permission being granted (passive form)

Table 5.2 Extended units of meaning and their semantic prosodies

Verb	Extended units of meaning	Semantic prosodies
/sǝɔn/	[person who teaches] /sǝɔn hây/ [person who is taught] [verb] ([object/adverb])	Someone being taught to do something
/càtkaan/	<p>[(expression of request)] [speaker-self reference] /càtkaan ʔeeŋ/ [(expression of request)]</p> <p>[person] /càtkaan hây/ </p>	<p>An offer to manage or deal with something on behalf of the interlocutor</p> <p>Arranging for something to happen</p>
/hǎay/	[person] /hǎay/ [verb expressing (negative) physical state or emotion]	Unpleasant emotion or physical state coming to an end
/baŋkháp/	<p>[person/thing that forces] /baŋkháp (mây) hây/ [person/thing that is forced] [verb] ([object/adverb])</p> <p>[person/thing that forces] /baŋkháp/ [person/thing that is forced] /(mây) hây/ [verb] ([object/adverb])</p> <p>[person/thing that is forced] /thùuk/ [person/thing that forces] /baŋkháp hây/ [verb] ([object/adverb]) /doon/</p> <p>[regulation] /mii phǝn baŋkháp/ /mii phǝn baŋkháp cháy/ /mii phǝn cháy baŋkháp/ [regulation] /cháy baŋkháp/ [regulation] /baŋkháp cháy/</p>	<p>Someone or something making someone (else) do something or something happen (active form)</p> <p>Someone or something making someone (else) do something or something happen (active form)</p> <p>Someone or something making someone (else) do something or something happen (passive form)</p> <p>The coming into effect of regulation</p> <p>Enforcement of regulation</p>
/duuleɛ/	[person] /duuleɛ (/mây/) hây/ [another person/thing] [verb] ([object/adverb])	Assurance that something that should or should not happen does indeed happen or not happen
/còp/	[something unpleasant] /mây còp sîn/ /mây mii wan còp sîn/ /mây rúucàk còp sîn/	Something unpleasant continuing indefinitely

Table 5.2 Extended units of meaning and their semantic prosodies

Many of these suggested extended units of meaning have the causative marker /hây/ as part of their core. In all cases where /hây/ forms part of the core of an extended unit of meaning, the pragmatic function of the unit is causation in some way. All thus operate as what is called a periphrastic causative in traditional grammar or a causative serialisation in Thai grammar (see section 1.3.3.2). However, the precise causative meaning depends on what verb /hây/ forms the core with. For example, the extended unit of meaning whose core is /bạnháp hây/ has the pragmatic function of expressing that someone *makes* someone else do something. On the other hand, the extended unit of meaning whose core is /duulɛɛ (mây) hây/ has the pragmatic function of expressing that someone makes something happen or not happen in accordance with what *should* be the case. As well as these extended units of meaning, the unit whose core is /càp/ also has the pragmatic function of causation, even though it does not include /hây/.

5.6.3 Extended units of meaning reconsidered from the polarity-oriented perspective

In an earlier chapter, we considered Tognini-Bonelli's analysis of *in the case of* (in section 2.7.7), and saw that this opens the way for some possible criticisms of the EUM-oriented approach. Particularly, it was argued that the semantic prosody of "introduction of specificity" for *in the case of* could be viewed, especially from the perspective of scholars who employ the polarity-oriented approach, as unusual or even wrong, as it does not incorporate any expression of attitude or evaluation. Similarly, the argument was presented that the "introduction of specificity" could in fact be seen as simply a paraphrase of the meaning of the core of the unit (ie. *case*), and therefore that this type of analysis of semantic prosody is pointless.

Being based on the Sinclairian perspective, like Tognini-Bonelli's analysis, my analysis above might invite the same criticism. That is, it might be objected that some of the semantic prosodies that I identified and presented in Table 5.2 are unusual or not even semantic prosodies at all, because they do not incorporate any expression of attitudinal or evaluative meaning. Examples would be the semantic prosody of "someone starting to do something" or the semantic prosody of "the coming into effect of regulation". In the discussion of Tognini-Bonelli's analysis of *in the case of*, I argued that, within the Sinclairian approach, semantic prosody can be any pragmatic function or meaning rather than being limited to expressions of positive or negative attitudinal or evaluative meaning, and that Tognini-Bonelli's analysis thus is exactly in line with Sinclair's approach. Here, I likewise maintain that the semantic prosodies that I identified are not unusual or wrong from the perspective of the EUM-oriented approach, even though scholars operating within the polarity-oriented approach might consider such an analysis to have strayed away from their definition of semantic prosody.

However, aside from these issues, some of the semantic prosodies that I identified remain vulnerable to the second criticism mentioned above. It might be objected that the semantic prosodies identified are a paraphrase of the meaning of the core of the unit, as with Tognini-Bonelli's analysis of *in the case of* where the semantic prosody of "introduction of specificity" could be seen as a paraphrase of *case*, which can mean 'a specific example'. Despite the fact that I adopt the same method and notion of semantic prosody as Tognini-Bonelli, I would argue that my analysis is less vulnerable to such a criticism than Tognini-Bonelli or even Sinclair's. I earlier contended that for a pattern to be identified as an extended unit of meaning, it must have a clear pragmatic function that

is distinct from a combination of the meanings considered in isolation, either literal or metaphorical, of the individual parts of the pattern (see section 3.4.2). Given that I adopt this criterion, my analysis should, unlike Tognini-Bonelli's, be less susceptible to the criticism in question.

To illustrate this point, let us now consider one example of a semantic prosody that I identified which might be seen as simply a paraphrase of the core of the unit, and my defence for it. It might be argued that the semantic prosody of “permission being granted” for the unit whose core is /ʔànúyâat hây/ is close to the meaning of /ʔànúyâat/, which means ‘allow’ or ‘permit’, and thus could be seen as just a paraphrase of the core. I would argue that although obviously the meaning of /ʔànúyâat/ does contribute to the semantic prosody of the unit, the overall pragmatic function that I identified is not the sum of the meanings of /ʔànúyâat/ and /hây/. /hây/ on its own means ‘give’. However, when preceded by /ʔànúyâat/ and used in the sequence /ʔànúyâat hây/, /hây/ drops the meaning of ‘give’, and rather functions as a causative marker. Therefore, the overall pragmatic meaning that the sequence expresses is ‘someone allows someone else to do something’. This pragmatic function is distinct from the combination of the meanings of /ʔànúyâat/ and /hây/ considered in isolation, and is thus not (or at least, is more than just) a paraphrase of the two items. This argument also applies to other extended units of meaning involving /hây/ as a causative marker.

I have so far established that, given that one of my criteria for a pattern to be considered as an extended unit of meaning is for its pragmatic function to be distinct from a combination of the meanings considered in isolation of the parts of the pattern, the semantic prosodies that I identified should be seen as less vulnerable to the criticism of

paraphrasing. I would further argue that the conservative approach that I adhere to also in fact helps to insulate my analysis from such a criticism (see section 3.4.2). Throughout the analysis, I identified a pattern as an extended unit of meaning only when was I very confident that the pattern in question met my criteria (see section 3.4.2). In cases where I was not sure, I ruled the pattern out from being an extended unit of meaning. For example, in my analysis of /hǎay/ ‘recover’, I ruled out the pattern /hǎay/ + /càak/ + /ròok/ from being an extended unit of meaning, despite the fact that the pattern is rather fixed and has its own specific colligation and semantic preference. The reason for this was that the pragmatic meaning of the sequence of “recovery from particular disease” is too close to the sum of the meanings of /hǎay/ ‘recover’ + /càak/ ‘from’ + /ròok/ ‘disease’ (see section 5.5.16). By contrast, scholars who do not employ the conservative approach or are less conservative than me, such as Tognini-Bonelli, (as evident from her analysis of *in the case of*), or Sinclair himself, might wish to consider this pattern as an extended unit of meaning.

So far in this section, then, we have considered two possible problems that scholars adopting the polarity-oriented approach might find with the EUM-oriented approach and particularly with the semantic prosodies that I identified in my analysis. It has been argued that attitudinal or evaluative meaning is central to the concept of semantic prosody, especially from the polarity-oriented perspective, although it has also been established that from the EUM-oriented perspective, semantic prosody can be any pragmatic function and is not confined to expressions of attitude or evaluation (see section 2.7.7). Given that scholars operating within the polarity-oriented approach consider attitudinal or evaluative meaning to be the defining feature of semantic prosody,

and that because of this, they might find problems with my analysis, I will now step out of the EUM-oriented approach and consider the semantic prosodies that I identified from the polarity-oriented approach, and attempt to characterise their positive or negative evaluation. For some of the words analysed, no additional analysis was required to accomplish this. For example, in some cases, such as the case of the unit whose core is /càp dây/ and the case of the unit whose core is /hăay/, I identified the evaluation based on the pragmatic meanings that I had already identified in the course of applying the EUM-oriented approach. However, where this was not possible, I undertook some further analysis to identify the evaluation of the semantic prosodies, by looking at the words or phrases that occur in the immediate context of the extended units of meaning within my concordance data. Table 5.3 lists the core of each extended unit of meaning, its semantic prosody, and its evaluation.

Extended unit of meaning (core)	Semantic prosody	Evaluation
/pràkàat hây/	Announcement in which a person, particularly one with authority, asserts that he/she will make something happen	Neutral
/càp muu/	Cooperation in doing something between two (groups of) people or organisations	Positive
/càp/	Someone making someone else do something	Neutral
/càp dâ/	(Bad) secret being revealed	Negative
/hă/	Someone starting to do something	Neutral
/yɔɔm hây/	Someone allowing someone else to do something or something to happen	Positive
/yók tuayàaŋ chên/	Linking a statement of some topic or phenomenon to an example of that topic or phenomenon	Neutral
/khûapkhum tua/	Someone being confined	Negative
/rɔɔ kaan kamnòt thòot/	Suspending determination of punishment in legal contexts	Neutral
/rɔɔ kaan loŋthòot/	Suspending a sentence in legal contexts	Neutral
/rɔɔ loŋ ʔaayaa/	Suspending a sentence in legal contexts	Neutral
/ʔànúyâat hây/	Permission being granted	Positive
/sɔɔn hây/	Someone being taught to do something	Positive
/càtkaan ʔeeŋ/	An offer to manage or deal with something on behalf of the interlocutor	Positive
/càtkaan hây/	Arranging for something to happen	Positive
/hăy/	Unpleasant emotion or physical state coming to an end	Positive
/baŋkháp (mây) hây/	Someone or something making someone (else) do something or something happen	Negative
/mii phôn baŋkháp/ /mii phôn baŋkháp cháy/ /mii phôn cháy baŋkháp/	The coming into effect of regulation	Neutral
/cháy baŋkháp/ /baŋkháp cháy/	Enforcement of regulation	Neutral
/duuleɛ (mây) hây/	Assurance that something that should or should not happen does indeed happen or not happen	Positive
/mây còp sîn/ /mây mii wan còp sîn/ /mây rúu càk còp sîn/	Something unpleasant continuing indefinitely	Negative

Table 5.3 Extended units of meaning, their semantic prosodies and evaluation

Some of the semantic prosodies listed in Table 5.3 are arguably associated with the expression of positive or negative evaluation. Others do not express positive or negative evaluation, that is, they are neutral in terms of evaluation.

The semantic prosodies of the units whose cores are /càtkaan ʔeeŋ/, /càtkaan hây/, /hăay/ and /duuleɛ (mây) hây/ apparently incorporate expressions of positive evaluation. The semantic prosodies of the units whose core are /càp dây/, /khûapkhum tua/, and /mây còp sîn/ (and its variants), by contrast, clearly incorporate expressions of negative evaluation. We see, for example, that the semantic prosody of “unpleasant emotion or physical state coming to an end” for /hăay/ is straightforwardly positive, whereas the semantic prosody of “something unpleasant continuing indefinitely” for /mây còp sîn/ (and its variants) is negative.

Although the evaluation of the semantic prosody of “(bad) secret being revealed” for /càp dây/, and that of the semantic prosody of “someone being confined” for /khûapkhum tua/, are negative, it might be argued that the evaluation of such semantic prosodies is also contingent upon whose point of view is being recounted. Hunston (2007: 256) discusses the issue of evaluation and point of view, arguing that they are “essentially linked”. An example she gives to illustrate this point is the use of the predicative adjective *persistent* in a news report about the stealing of a cardboard figure from a store. In the store manager’s accounts of the accident, the manager comments that “There were three young guys out to get him [the figure] and they were very persistent.” Hunston argues that from the viewpoint of the thieves, *persistent* is positive, as their being persistent enables them to steal the figure. But *persistent* is negative from the viewpoint of the store manager, because the thieves’ being persistent prevents him/her

from protecting the figure from being stolen. Similarly, it might therefore be argued that the disclosure of a (bad) secret is negative from the point of view of a person whose (bad) secret is revealed, but positive from the point of view of a person who benefits from the (bad) secret being revealed. Likewise, confinement is arguably negative from the viewpoint of a person who is confined, often a person who breaks the law, because he/she fails to escape but ends up being arrested and confined. It is, by contrast, positive from the viewpoint of a person who confines, often an officer, and depending on the offense, perhaps also from the perspective of society at large.

The semantic prosodies of the units whose cores are /càp mɯɯ/, /yɔɔm hây/, /ʔànúyâat hây/ and /sǎɔn hây/ arguably incorporate expressions of positive evaluation. Cooperation is generally seen as positive. Similarly, the state of being allowed to do something and of being taught to do something could arguably be seen as positive, especially from the point of view of the person who receives permission or is taught.

The semantic prosody of the unit whose core is /baŋkháp (mây) hây/ is, by contrast, arguably incorporates expressions of negative evaluation. /baŋkháp/ on its own means ‘force’. But the sequence of /baŋkháp (mây) hây/ has the pragmatic function of expressing that someone or something makes someone (else) do something or something happen. This function seems on the surface to be neutral in terms of evaluation. Looking at the actions the causee, whether human or non-human, is made to perform, I also found that most (40 out of 215) are neutral in terms of evaluation. That said, the overall meaning the sequence expresses incorporate the sense of involuntariness, especially when the causee is human. Thus, I would argue that even though the actions caused may be in themselves evaluatively neutral, the overall evaluation of the unit is negative.

Unlike the semantic prosody of unit whose core is /baŋkháp (mây) hây/, which I argued above to display negative evaluation, the semantic prosody of the unit whose core is /càp/ does not seem to display the same negative evaluation, despite the fact that it has much the same pragmatic function of expressing that someone makes someone else do something. Looking at all 15 instances of /càp/ used in contexts of causation, I found 10 instances where the action the causee is made to perform is evaluatively neutral, 4 where the action is evaluatively negative, and 1 where the action is evaluatively positive. Arguably then, unlike the semantic prosody of the unit whose core is /baŋkháp (mây) hây/, the semantic prosody of this unit displays neutral evaluation.

The evaluation of semantic prosodies of “someone starting to do something” for /hăn/ is arguably neutral. The evaluation of “someone starting to do something” is in itself neutral. Examining the evaluation of the actions the subject starts to do, I also found that most (237 out of 374) are neutral. There are 100 instances where the action is positive, and 37 where the action is negative. Thus, I would argue that the evaluation of the semantic prosody of “someone starting to do something” is neutral in evaluation.

As with the evaluation of the semantic prosody of “someone starting to do something”, the evaluation of the semantic prosody of “announcement in which a person, particularly one with authority, asserts that he/she will make something happen” for /pràkàat hây/ is in itself neutral. Investigating the actions that the announcement makes happen, I also discovered that most (27 out of 30) are evaluatively neutral. I thus consider this semantic prosody to be neutral in terms of evaluation.

For the remaining six semantic prosodies which are the semantic prosodies of the units whose cores are /yók tuayàaŋ chên/, /rɔɔ kaan kamnòt thòt/, /rɔɔ kaan loŋthòt/,

/rɔɔ loŋ ʔaayaa/, /mii phõn baŋkháp/ (and its variants), and /cháy baŋkháp/ (and its variant), I argue that they are all neutral in terms of evaluation. While /yók tuayàaŋ chên/ means ‘for example’ and is frequently used in *academic* (see section 5.5.7), the other four units are used in legal contexts and not associated with evaluation (see sections 5.5.12 and 5.5.17).

We have seen from the above analysis that whereas some semantic prosodies that are identified based on the EUM-oriented approach can be seen as incorporating expressions of positive or negative evaluation, others cannot. Given that scholars operating from the polarity-oriented perspective consider evaluative meaning to be the central feature of semantic prosody, they may find the semantic prosodies that do not involve aspects of evaluation uninteresting or even wrong. However, from the EUM-oriented approach, such semantic prosodies merit as much attention as those associated with positive or negative evaluation, as both types express a pragmatic function of the extended units of meaning identified. Within the Sinclairian approach, as has been argued throughout the thesis, semantic prosody can be any pragmatic function and is not limited to expressions of positive or negative evaluation. We see, for example, that the unit whose core is /yók tuayàaŋ chên/ has a pragmatic function of structuring information in the text or discourse, and this function is not in any particular way associated with evaluation.

5.6.4 Variation in use of single-word units of meaning

We have seen variation across genres when the verb is used as a single-word unit of meaning. Mostly, the variation exists in terms of the frequency of occurrence of different patterns with these verbs in the four genres. For example, /khûapkhum tuaʔeeŋ/ and /khûapkhum tonʔeeŋ/ are prevalent only in *fiction*. Similarly, the sequence of /càtkaan/ plus a human object occurs frequently only in *fiction* and *non-academic*, but rarely in *newspaper*, and absent in *academic*, whereas the sequence /càtkaan/ plus a food item occurs only in *fiction*, and not in the other three genres. This underlines the point that different co-occurrence patterns of the same verb may have different frequency distribution across genres. We have also seen some cases where a given word has different meanings, and one particular meaning dominates a particular genre, while another meaning dominates another genre. /yók/, for instance, has multiple senses; two of them are ‘lift’ and ‘refer’. While in *academic*, the meaning ‘refer’ is the most prevalent, in *fiction*, *newspaper*, and *non-academic*, the meaning ‘lift’ is the most frequent.

5.6.5 Variation in use of extended units of meaning and semantic prosodies

So far I have discussed a range of interesting issues that have come up recurrently throughout the analysis of the 19 verbs. Let me now return to the main aim of this chapter, which is to answer the research question of what variation in semantic prosodies can be identified for Thai words across genres. To answer this research question, I will consider the extended units of meaning identified for each verb and their semantic prosodies across genres.

We have seen throughout the analysis that the use of the extended units of meaning identified for each verb varies across genres. We have established that semantic prosody is definitionally linked to an extended unit of meaning, as it is a compulsory element of the unit (see section 2.3.2.1). Thus, logically, variation in the use of the extended units of meaning across genres also implies variation in semantic prosodies across genres. In the analysis of the 19 verbs, we have seen that while some extended units of meaning occur only in specific genres, others occur across all four genres, but vary in terms of frequency of occurrence. Table 5.4 gives the proposed extended units of meaning and the frequency of occurrence of each unit across all four genres. The numbers presented in the table are absolute frequencies. These figures are directly comparable with one another, because they are all based on exactly 200 random examples of each verb in each genre. It is not necessary to present a relative frequency in this context. Moreover, because the counts presented are based only on the samples, a relative frequency per million words would be misleading here.

Extended unit of meaning (core)	Academic	Fiction	Newspaper	Non- academic
/pràkàat hây/	3	9	10	8
/càp muu/	1	-	17	1
/càp/	1	6	4	4
/càp dâ/	1	12	5	5
/hăn/	142	8	123	101
/yoom hây/	38	17	22	18
/yók tuayàaη chên/	20	-	2	12
/khûapkhum tua/	6	2	34	1
/rɔɔ kaan kamnòt thòt/	4	-	-	-
/rɔɔ kaan loηthòt/	5	-	-	-
/rɔɔ loη ʔaayaa/	2	-	3	-
/ʔànúyâat hây/	50	58	55	69
/sɔɔn hây/	18	20	16	35
/càtkaan ʔeeη/	-	22	1	-
/càtkaan hây/	3	22	4	3
/hăy/	9	31	20	19
/baηkháp (mây) hây/	24	88	40	63
/mii phôn baηkháp/ /mii phôn baηkháp cháy/ /mii phôn cháy baηkháp/	16	-	33	6
/cháy baηkháp/ /baηkháp cháy/	38	-	34	10
/duuleε (mây) hây/	5	1	5	5
/mây còp sîn/ /mây mii wan còp sîn/ /mây rúu càk còp sîn/	1	4	2	2

Table 5.4 Extended units of meaning and their frequency of occurrence across genres

Table 5.4 shows that some extended units of meaning occur in specific genres. Others occur across all four genres but vary in terms of frequency of occurrence. In what follows, I will first discuss the extended units of meaning that are genre-specific, and then those that are present across all four genres.

Of the 21 extended units of meaning, 8 are restricted to one, two, or three of the four genres. The units whose cores are /rɔɔ kaan kamnòt thòt/ and /rɔɔ kaan loηthòt/ are restricted to *academic*. This is not unexpected considering the fact that they are technical legal terms mostly used in legal contexts. They are therefore likely to be found

in legal academic texts rather than in fiction or newspapers. Looking at the metadata of the texts where they occur, I found that these texts are indeed drawn from the legal textbooks.

Of the remaining six restricted units, four occur in *academic*, *newspaper*, and *non-academic*, but are absent in *fiction*. These are the units whose cores are /càp mɯɯ/, /yók tuayàŋ chên/, /mii phõn bạŋkháp/ (and variants), and /cháy bạŋkháp/ (and variant).

The unit whose core is /càp mɯɯ/ occurs only once in each of *academic* and *non-academic*, but as many as 17 times in *newspaper*. Its frequent occurrence in *newspaper* can be explained by the fact that the expression of a situation where two organisations are cooperating, which is the semantic prosody of the unit, is a topic that seems likely to be prominent in news journalism. We are therefore less likely to find this extended unit, with its overtones of news reportage, in *academic*, *non-academic*, and especially *fiction*, where a writer narrates a story rather than reporting current news.

The unit whose core is /yók tuayàŋ chên/ is especially frequent in *academic*. It is also quite frequent in *non-academic*. There are just two instances in *newspaper*, and none in *fiction*. This is not unexpected considering the unit's semantic prosody of linking a statement of a phenomenon to an example of that phenomenon. Giving an example to elaborate a definition or a point being made is a common rhetorical requirement in academic writing, but much less so in journalism or fiction.

The unit whose core is /mii phõn bạŋkháp/ (and variants) is especially frequent in *newspaper*. The unit whose core is /cháy bạŋkháp/ (and variant) is particularly frequent in *newspaper* but also in *academic*. Both units are technical legal terms used in legal contexts, and are thus likely to be found in legal academic texts. They are also likely to

be found in *newspaper*, as the coming into effect of regulation and enforcement of regulation, which are the units' semantic prosodies, can also be a topic prominent in political news.

The remaining two extended units of meaning that are found in fewer than four genres are /rɔɔ loŋ ʔaayaa/ and /càtkaan ʔeeŋ/. The unit whose core is /rɔɔ loŋ ʔaayaa/ occurs a few times only in *academic* and *newspaper*. They are again technical legal terms. The unit whose core is /càtkaan ʔeeŋ/ is exceptionally frequent in *fiction*, but absent in *academic* and *non-academic*. It occurs only once in *newspaper*. The reason why this unit occurs frequently in *fiction* may be because its pragmatic function, of expressing an offer to arrange something on behalf of someone else, is an interpersonal one, and we are likely to find the interpersonal interaction in fictional dialogues.

The remaining 13 extended units of meaning are found in all four genres but vary in terms of frequency of occurrence. Of these 13 units, there are cases where that variation is arguably considerable, in that the frequency of occurrence is exceptionally high or low in one specific genre relative to the other three. But there are also cases where that variation in frequency across the genres seems to be limited.

It is interesting that the unit whose core is /hǎn/ is frequent everywhere except in *fiction*. It occurs only eight times in *fiction*, but more than a hundred times in each of the other three genres. In *academic*, *newspaper* and *non-academic*, /hǎn/ is most frequently used in its metaphorical sense, forming part of the extended unit of meaning that has the semantic prosody of expressing someone starting to do something. In *fiction*, by contrast, /hǎn/ is mostly used in its literal sense as a single-word unit of meaning, i.e. meaning 'turn', the physical, corporeal sense. This more frequent use of literal /hǎn/ can probably

be explained by the fact that it is more common to describe the physical actions of characters in a story than to express the inception of an action, due to the nature of a narrative as a sequence of (complete) events.

The variation is also arguably high with the extended units of meaning whose cores are /khûapkhum tua/ and /càtkaan hây/. The unit whose core is /khûapkhum tua/ is frequent in *newspaper*, but not in the other three genres. The unit's frequency in *newspaper* is probably due to the fact that the unit's semantic prosody, of expressing that someone is confined (e.g. an imprisoned criminal), is likely to be a prominent topic in news reportage. The unit whose core is /càtkaan hây/ is common in *fiction*: there are just a few instances in each of *academic*, *newspaper*, and *non-academic*, but as many as 22 in *fiction*. Its frequency in *fiction* is probably due to the fact that its pragmatic function, of expressing an arrangement for something to happen, is interpersonal, and thus likely to appear in fictional dialogues (in the same way as /càtkaan ʔeeŋ/, discussed above).

The level of variation across genres is relatively low for the remaining ten extended units of meaning. The frequency of the unit whose core is /ʔànúyâat hây/, for example, ranges between about 50 and 70, with the lowest frequency of 50 in *academic* and the highest frequency of 69 in *non-academic*. The unit whose core is /mây còp sîn/ (and variants) occurs once in *academic*, twice in each of *newspaper* and *non-academic*, and four times in *fiction*.

One interesting point is observed with the units whose cores are /càp/, /càp dây/, /hăay/, and /banháp (mây) hây/. These appear to share a particular pattern of variation, in that the frequency is highest in *fiction* and lowest in *academic*, whereas the frequencies in *newspaper* and *non-academic* are reasonably similar to that in *fiction*. For instance, the

unit whose core is /càp/ occurs four times in each of *newspaper* and *non-academic*, six times in *fiction*, but only once in *academic*. This can probably be explained by the fact that the units' semantic prosodies are interpersonal and, as already argued, thus may be dispreferred in academic writing, especially relative to fiction.

This last is, in fact, a finding that has emerged repeatedly in this discussion: a contrast between *academic* and *fiction*, where units that occur frequently in one of these two genres often tend to occur infrequently in the other. For example, the unit whose core is /hǎn/ occurs most frequently in *academic*, but least frequently in *fiction*. Similarly, the units whose cores are /yók tuayàaŋ chên/ and /cháy baŋkháp/ (and variant) are most frequent in *academic*, but are absent in *fiction*. By contrast, the units whose cores are /càtkaan ʔeeŋ/ and /càtkaan hây/ are most common in *fiction*, but least common in *academic*. This pattern is arguably due to the substantial difference in textual function between *academic* and *fiction*. As Biber *et al* (1998: 149-152) argue in their study of variation across registers in English, academic prose is inclined towards informational production, but fiction is inclined towards involved production, which is concerned more about interpersonal interaction and personal feelings and concerns, rather than giving information. This appears to explain my observations in the Thai data. I have earlier argued, for example, that the units whose cores are /càp/, /càp dây/, /càtkaan ʔeeŋ/ /càtkaan hây/, /hǎay/, and /baŋkháp (mây) hây/ are frequent in *fiction* because their pragmatic functions (of expressing an offer) reflect interpersonal interaction. The unit whose core is /yók tuayàaŋ chên/, on the other hand, is common in *academic*, and this, I have argued, is because it is a common rhetorical requirement in academic writing to give an example to elaborate the point being made.

5.7 Concluding remarks

In this chapter, I have investigated variation in semantic prosodies across genres. To do so, I have analysed the use of 19 Thai verbs across four different genres: *academic*, *fiction*, *newspaper*, and *non-academic*. The results of the analysis show that almost all the verbs under study, except /hǎn/ ‘turn’, are used most frequently as a single-word unit of meaning. That said, some verbs do form part of the extended units of meaning that have a clear semantic prosody or pragmatic function in the Sinclairian sense. We have seen variation across genres not only when the verbs are used as single-word units of meaning, but also as when it forms part of the extended units of meaning. The level of variation across genres in the use of the extended units of meaning, which implies variation in semantic prosodies, is considerable with some units, but limited with others. Specifically, we have seen a notable contrast in the use of the extended units of meaning and semantic prosodies between *academic* and *fiction*, in that the units that occur frequently in one of these two genres are often likely to occur infrequently in the other.

It is also worth noting that although the number of the extended units of meaning that I identified may seem small, to identify these extended units of meaning and their semantic prosodies and the associated cross-genre variation, I had to manually investigate 15,200 concordance lines. It was a difficult and tremendously time-consuming task. However, the results are well worth the effort spent, as it allows us to have a better understanding of variation in semantic prosodies across genres in the Thai language.

Chapter 6 – Analysis of semantic prosodies of translation-equivalents

6.1 Introduction

In this chapter, I address the third research question: to what extent are the semantic prosodies of translation-equivalents in Thai and English similar or different? I will first present the approach that will be adopted in the analysis in section 6.2. I then go on to provide a discussion of the concept of *prosodic strength* in section 6.3, which will become important to the analysis. In section 6.4, I will present the criteria for the selection of translation-equivalent pairs, and the translation-equivalent pairs whose semantic prosody will be studied. Then in section 6.5, I will present the results of the analysis, including a list of collocates for each translation-equivalent pair. This section also includes a discussion of some methodological issues which arose in the course of conducting the analysis to arrive at the semantic prosodies of each translation-equivalent pair. In section 6.6, I give a detailed discussion of the analysis results.

6.2 Approach to the analysis

The aim of the third research question is to investigate to what extent the semantic prosodies of translation-equivalents in Thai and English are similar or different. In order to identify the semantic prosodies of the translation-equivalents in question, the polarity-oriented approach was employed (see section 3.4.1). This approach is suitable to address this research question because the goal is to examine whether or not translation-equivalents in Thai and English have similar tendencies to appear in positively or negatively evaluated contexts. Practical work on semantic prosody in translation or

contrastive studies has indeed favoured the polarity-oriented approach, as it reveals a positive or negative evaluative potential of a word, over a Sinclair-style phraseological approach, as was in fact evident in the literature review (see section 2.7). Moreover, in a context where I wish to investigate a large number of translation-equivalent pairs, this approach is more practical than the EUM-oriented approach, which requires much more research time and space in the thesis to be devoted to each word.

As with the collocate analysis method adopted in Chapter 4, Log Ratio was used as the measure of collocational strength. Only items with Log Ratio score of three or more than three that occur in at least five different texts were considered collocates of a given node. Semantic prosody of the translation-equivalent pairs is confined to being positive or negative (or neutral), according to the rule of thumb discussed in section 3.4.1.1. A collocate marked in bold is negatively evaluated, and a collocate marked by an underline is positively evaluated. Collocates marked by neither are neutrally evaluated.

6.3 Prosodic strength

As well as investigating the semantic prosodies of the translation-equivalents, I looked at the strength of the prosody in each case. It is worth investigating prosodic strength in this context, as prosodic strength has been argued to be an indicator of degrees of equivalence (Wei and Li 2013: 109) and has been examined by previous studies of semantic prosody across languages, such as Wei and Li (2013) and Ebeling (2013).

Semantic prosody, according to the polarity-oriented approach, is a preference for positive or negative collocating words. *Prosodic strength* refers to the degree of this preference. Wei and Li (2013: 109) mean by semantic prosodic strength “the node item’s

tendency and attraction for a particular type of attitudinal meaning.” They argue that, as well as the polarity of semantic prosody, prosodic strength is another important indicator of “degrees of equivalence between translation equivalents” (Wei and Li 2013: 109). In cases where the words in a translation pair display the same polarity of semantic prosody, the closer the prosodic strength of the words in the pair, the greater the degree of equivalence. Wei and Li’s notion of prosodic strength is arguably similar to Partington’s (2004: 153) idea of *degrees* of semantic prosody, in which words that have the same prosody can have a more positive or a less positive prosody, or a more negative or a less negative prosody. However, Partington does not discuss the notion in the context of translation equivalence (see section 2.3.4.1). Like Wei and Li, Ebeling (2013: 11) touches upon the concept of degrees of semantic prosody in the context of contrastive studies. Investigating the translation-equivalents of the lemma *cause* in Norwegian, Ebeling finds that there are no Norwegian translation-equivalents that have the same degree of negative semantic prosody as *cause*. Ebeling (2013: 11) therefore goes on to argue that “there are clear counterparts to *cause* in Norwegian, i.e. there are correspondences (as shown in the corpus data), but no real equivalent of *cause*.” This argument suggests that, for Ebeling, even when the words in a translation pair have the same semantic prosody, if the degree to which the two words possess that prosody is different, they are merely counterparts, and not “real” equivalents.

Although Wei and Li, Partington, and Ebeling all discuss the idea of prosodic strength, only Wei and Li (2013: 109) propose a formula for calculating prosodic strength. In their study of the semantic prosodies of translation-equivalents in Chinese and English, they examine 100 concordance lines of each node to extract collocates or

see what type of attitudinal meaning, positive or negative, is associated with the node. Wei and Li (2013: 108-109) quantify prosodic strength in terms of a fraction of the number of tokens of the node examined. Therefore, they argue, for example, that the lemma *spring up* has a positive prosodic strength of 0.63, because out of 100 concordance lines, there are 63 cases where *spring up* is associated with a positive attitudinal meaning.

In my analysis, I will not adopt Wei and Li's formula for calculating prosodic strength. Instead, I will consider the ratio of positive collocate types to negative collocate types in cases where a word has a positive semantic prosody, or the ratio of negative collocate types to positive collocate types in cases where a word has a negative semantic prosody. The reason for this is based on my method, as was explained in section 3.4.1.1. When I assign a positive or negative semantic prosody to a word, I primarily consider whether the number of positive collocate types is greater than the number of negative collocate types, or vice versa. That is, in my method, the number of positive collocate types has to be at least three times bigger than that of negative collocate types for a word to be considered to have a positive prosody. In the same way, the number of negative collocate types has to be three times or more bigger than that of positive collocate types for a word to be argued to have a negative prosody. I do not take into account the number of neutral collocate types, unless it accounts for 70% or more of the total number of collocate types. (See section 3.4.1.1 for the criteria for assigning a positive, negative, or neutral semantic prosody.) In the context of a criterion based on a ratio of types, it is more reasonable to calculate prosodic strength by considering the ratio of positive collocate types to negative collocate types, and vice versa, than by quantifying it in terms

of a fraction of the number of the tokens of the node examined. However, for clarity, I will also present percentages of collocate types alongside the actual number of collocate types.

6.4 Selection of translation-equivalent pairs

6.4.1 Selection of Thai words

This chapter primarily focuses on exploring the semantic prosodies of translation-equivalent pairs that have not clearly been the object of earlier research. One translation-equivalent pair whose English member has been extensively studied was explored in Chapter 4 (*/kəʊzə/* ‘cause’). The Thai words under study here were, by contrast, not selected on the basis that their translation-equivalents in English have been studied. Rather, the criteria for Thai word selection were set as follows:

1. Only words with overall corpus frequency between 5,000 and 7,500 were included.
2. Only complex (i.e. polymorphemic) words were considered. This is due to tokenisation issues. There are quite a lot of instances in the corpus where what is tokenised as a simplex word is actually part of a complex word. Therefore, it is likely that statistical collocates obtained from the automated analysis of a simplex word are in fact collocates of one or other complex word of which it may form a part. It is much less unlikely, on the other hand, that this problem will arise for collocates of complex words.

3. Only nouns were analysed. In Chapter 4 and Chapter 5, the words under study were all verbs. It is thus worth investigating another part of speech.

6.4.2 Selection of English translation-equivalents

The English translation-equivalents were identified using two bilingual dictionaries, *SE-ED's Modern Thai-English Dictionary* (Thiengburanatham 2001) and *Thai English German Dictionary* (Rohrer 2007). These are the best bilingual dictionaries I could find on the market. For reasons of space and time, only the first equivalent listed in either of the two dictionaries was analysed. In many cases, the first equivalent listed in *SE-ED's Modern Thai-English Dictionary* is the same as the one listed in *Thai English German Dictionary*. In cases where the first equivalents listed in the two dictionaries are different, I used my intuition as a native speaker of Thai to select the equivalent I considered closest in terms of meaning to the Thai word. One example is /chaawbâan/. The English equivalent provided by *SE-ED's Modern Thai-English Dictionary* is *rustics*, whereas the one provided by *Thai English German Dictionary* is *villager*. Even though both *rustics* and *villager* mean people in the countryside, I considered *villager* a better translation-equivalent of /chaawbâan/ than *rustics*. The reason was that the use of the term *rustics*, according to the *Macmillan Dictionary* of English, implies that you think that people from the countryside are rude and not intelligent, which is not implied by /chaawbâan/. Moreover, *rustics* is an archaism. For these reasons, *villager* was chosen over *rustics*. There were also cases where only one of the two dictionaries provides a translation-equivalent for the Thai word under investigation. In this situation, the translation-equivalent provided was automatically accepted.

There is one instance where the second translation, rather than the first translation, was selected for the analysis, namely /phâappháyon/. The first translation given by *SE-ED's Modern Thai-English Dictionary* is *cinema*, whereas that provided by *Thai English German Dictionary* is *motion picture*. I considered *motion picture* to be the better translation. However, there is not sufficient data to look at in the BNC; *motion picture* occurs only 85 times in the BNC. This is probably because *motion picture* is American English, and it is thus less likely to be a frequent means of expressing this meaning in a corpus of British English. In this case, I thus selected to analyse *film*, which is the second translation given by *Thai English German Dictionary*. *Film* is British English and occurs 12,559 times in the BNC.

56 Thai words meet the criteria listed in section 6.4.1. However, eight of these words, /phrácâw/, /nítikam/, /sàhàrát/, /khonnay/, /phráʔon/, /nùayṅaan/, /róyylá/, and /thīmaa/, were excluded. /phrácâw/ is a title of a king or a member of the royal family. There is no direct translation-equivalent in English. /nítikam/ is a legal technical term. Its translation-equivalent *juristic act* does not occur in the British National Corpus. /sàhàrát/ is generally a short term for ‘the United States of America’ or ‘the United Arab Emirates’. There is no single translation-equivalent in English. /khonnay/ is erroneously tokenised. /khonnay/, as a noun, means ‘insider’. However, in the corpus, most of the instances of /khonnay/ should have been tokenised as two words: /khon/ meaning ‘people’ and the preposition /nay/ meaning ‘in’. /phráʔon/ is an honorific personal pronoun for a king or a member of the royal family. There is no direct English translation-equivalent. Neither of the bilingual dictionaries I used provides a translation-equivalent for /nùayṅaan/, which I understand to mean *institute*. To be consistent with the

methodology, /nùayŋaan/ was therefore excluded. /rǒylá/ ‘percentage’ was excluded because almost all of its collocates are numbers, which I consider to be neutral for semantic prosody. Numbers can in fact be interesting to look at. For example, Hoey (2007: 33-42) examines *sixty* and its numerical form *60*. He finds that even though they are just numbers, they also have specific primings for collocations, colligations, and pragmatic functions. Nevertheless, this area of analysis is not the focus of my study. Finally, similarly to /khonnay/, /thîmaa/ is erroneously tokenised. /thîmaa/, as a noun, means ‘source’. However, in the corpus, there are a large number of instances where /thîmaa/ actually means ‘which come (from)’. Here, /thî/ is a relative pronoun and is followed by the verb /maa/ ‘come’. This is yet another example of an error in the automatic tokenisation. Table 6.1 lists the Thai words that meet the criteria and were analysed, their English translation-equivalents, their frequency in the TNC or BNC, and their frequency per million words.

	Thai word/ English translation- equivalent	Frequency	Frequency per million words
1.	/phǒnŋaan/ achievement	5,028 4,411	144.56 44.06
2.	/thǒŋthìn/ district	5,034 8,172	144.73 81.62
3.	/pràwàtsàat/ history	5,099 19,033	146.6 190.10
4.	/thoorásàp/ telephone	5,128 6,307	147.43 62.99
5.	/phûukhon/ people	5,130 101,707	147.49 1,015.86
6.	/phâappháyon/ film	5,156 12,559	148.24 125.44
7.	/pràsìthíphâap/ efficiency	5,156 3,626	148.24 36.22

Table 6.1 Selected Thai words, their English translation-equivalents, their frequency, and their frequency per million words

	Thai word/ English translation- equivalent	Frequency	Frequency per million words
8.	/ηâankhăy/ condition	5,199 22,708	149.47 226.81
9.	/phûuyày/ adult	5,351 7,453	153.84 74.44
10.	/sùkkhàphâap/ health	5,392 22,835	155.02 228.08
11.	/rátthàthammánuun/ constitution	5,396 3,916	155.14 39.11
12.	/pâwmăay/ target	5,430 8,514	156.11 85.04
13.	/chaaynùm/ young man	5,434 3,659	156.23 36.55
14.	/sìŋwêetlóm/ environment	5,473 13,661	157.35 136.45
15.	/pràsòpkaan/ experience	5,489 20,778	157.81 207.53
16.	/khwaamtôŋkaan/ desire	5,517 5,511	158.62 55.04
17.	/tàwantòk/ west	5,518 19,974	158.64 199.50
18.	/ʔitthíphon/ influence	5,556 9,056	159.74 90.45
19.	/chûamooŋ/ hour	5,559 25,200	159.82 251.70
20.	/phúunthăan/ foundation	5,571 4,947	160.17 49.41
21.	/sàthăanákkaan/ situation	5,574 17,501	160.25 174.80
22.	/khamtòp/ answer	5,575 10,830	160.28 108.17
23.	/raayláʔiat/ detail	5,617 16,600	161.49 165.80
24.	/phômêe/ parents	5,786 14,790	166.35 147.72
25.	/nêewkhít/ idea	5,808 28,598	166.98 285.64
26.	/ʔèekkàchon/ individual	5,880 12,834	169.00 128.19
27.	/naayókrátthàmontri/ prime minister	5,888 9,509	169.28 94.98

Table 6.1 Selected Thai words, their English translation-equivalents, their frequency, and their frequency per million words

	Thai word/ English translation- equivalent	Frequency	Frequency per million words
28.	/ʔèkkàsään/ document	5,943 8,627	170.86 86.17
29.	/nèewthaaŋ/ way	5,983 95,730	172.01 956.16
30.	/muaŋthay/ Thailand	6,013 872	172.88 8.71
31.	/lākthāan/ evidence	6,083 20,328	174.89 203.04
32.	/sāpsĭn/ property	6,086 15,910	174.97 158.91
33.	/khamsāŋ/ order	6,277 35,400	180.47 353.58
34.	/ŋóppràmaan/ budget	6,458 8,159	185.67 81.49
35.	/náamman/ oil	6,514 10,391	187.28 103.79
36.	/khâarâatchákaan/ civil servant	6,553 1,288	188.4 12.86
37.	/ʔoŋkoon/ company	6,620 55,128	190.33 550.62
38.	/cĭtcay/ mind	6,717 21,181	193.12 211.56
39.	/mâattràthāan/ standard	6,777 14,382	194.84 143.65
40.	/nákrian/ student	6,976 21,343	200.56 213.18
41.	/khwaamsāmâat/ ability	6,993 9,936	201.05 99.24
42.	/mùubāan/ village	7,046 12,954	202.57 129.39
43.	/sàthāanthĭi/ place	7,174 47,858	206.25 478.01
44.	/chaawbāan/ villager	7,211 912	207.32 9.11
45.	/kitcākam/ activity	7,254 22,213	208.55 221.87
46.	/bòtbàat/ role	7,271 19,937	209.04 199.13
47.	/khunnáphâap/ quality	7,383 17,537	212.26 175.16

Table 6.1 Selected Thai words, their English translation-equivalents, their frequency, and their frequency per million words

	Thai word/ English translation- equivalent	Frequency	Frequency per million words
48.	/kitcakaan/ business	7,490 36,205	215.34 361.62

Table 6.1 Selected Thai words, their English translation-equivalents, their frequency, and their frequency per million words

6.5 Results

In this section, I present my data in the same format that I used in Chapter 4, using the same notations. However, unlike in Chapter 4, I will simply present a list of collocates of each word. There will be no specific discussion after each word, due to the great difference in scope between the study in Chapter 4 and this study. Given this difference, however, I will discuss the methodological issues I encountered during the course of carrying out this second and much more extensive analysis. This will be followed by a table summarising the semantic prosody of each word, and some further discussion of the outcomes.

6.5.1 Lists of the collocates of each Thai word and its English translation-equivalent

6.5.1.1 /phǒnṅaan/ vs. *achievement*

The collocates of /phǒnṅaan/ ranked by Log Ratio score are as follows:

/boodeɛŋ/ ‘masterpiece’ /chínʔèek/ ‘eminent’ /diidèn/ ‘excellent’ /sàtròok/ ‘stroke’ /lǒklian/ ‘copy’ /sǎansǎn/ ‘create’ /wichaakaan/ ‘academic’ /sǐnámman/ ‘oil colour’ /pràməən/ ‘estimate’ /níthátsakaan/ ‘exhibition’ /tiiphim/ ‘publish’ /boo/ ‘bow’ /yǒtyiam/ ‘excellent’ /sǐnlápakam/ ‘art work’ /prākùat/ ‘contest’ /dòotdèn/ ‘distinguished’ /citrākam/ ‘painting’ /sǐnlápin/ ‘artist’ /diisaynəə/ ‘designer’ /chín/ ‘piece’ /róykyêew/ ‘prose’ /nákkhǎn/ ‘author’ /wícay/ ‘research’ /phâaphàay/ ‘photograph’ /sǐnlápà/ ‘art’ /phǎəyphrêɛ/ ‘disseminate’ /chomchəəy/ ‘praise’ /ʔòkbêep/ ‘design’

/návátàkam/ ‘innovation’ /châaṅphâap/ ‘photographer’ /raaṅwan/ ‘reward’ /rûaṅsân/ ‘short story’ /phûukamkap/ ‘director’ /wannákam/ ‘literary work’ /sǎn/ ‘select’ /líkkhàsit/ ‘copyright’ /phâakphuumcay/ ‘proud’ /chûunchom/ ‘pleased’ /pràcàk/ ‘evident, realise, empirical’ (person’s name) /phâapkhian/ ‘painting’ /yóonlǎṅ/ ‘retrospect’ /rûapruam/ ‘compile’ /phuumcay/ ‘proud’ /chíawchaan/ ‘expert’ /khunnáwút/ ‘qualification’ /pràdit/ ‘invent’ /lâasùt/ ‘latest’ /khàněṅ/ ‘branch’ /ʔâaṅʔiṅ/ ‘refer’ /choo/ ‘show’ /tittaam/ ‘follow’ /thópthuan/ ‘review’ /chûusǎṅ/ ‘fame’ /waarásāan/ ‘journal’ /fēechân/ ‘fashion’ /pràphan/ ‘compose’ /khát/ ‘select, copy’ /diisay/ ‘design’ /chánálǎot/ ‘winning’ /náváníyaay/ ‘novel’ /pràtyaa/ ‘philosophy’ /góɔp/ ‘golf’ /dèn/ ‘outstanding’ /chûudan/ ‘famous’ /kàwii/ ‘poet’ /khónkhwáa/ ‘study intensively’ /sàthâapàttàyakam/ ‘architecture’ /diaw/ ‘solo’ /thàlěṅ/ ‘declare’

The collocates of *achievement* ranked by Log Ratio score are as follows:

crowning stupendous remarkable pinnacle emulate outstanding educational notable **devalue** vocational technological proud sporting astonishing lasting academic excellence motivation artistic consolidate mathematics poetic record objective attainment lifetime extraordinary aspiration impressive tremendous celebrate modest architectural pride export glorious recognition praise literary goal fantastic reward profile pupil technical award **obstacle** considerable recognise supreme intellectual productivity sense satisfaction unique solid assess cultural positive certificate ambition significant great participation scientific recording qualification substantial thatcher emphasise acknowledge engineering personal

6.5.1.2 /thóoṅthin/ vs. *district*

The collocates of /thóoṅthin/ ranked by Log Ratio score are as follows:

/khôobanyàt/ ‘legal code’ local /sǒo thǒo/ (acronym for ‘Department of Local Administration’) /phuumpanyaa/ ‘intelligence’ /câwphánákjaan/ ‘official’ /thúrákandaan/ ‘far-off’ /caaríitpràpheenii/ ‘custom’ /phûubɔ̀ríhǎan/ ‘executive’ /ʔoṅkɔ̀onphátthánaaʔèkkàchon/ ‘non-governmental organisation’ /rátthàbaanklaaṅ/ ‘central government’ /khànòpthamniam/ ‘custom’ /sùanphuumípâak/ ‘provincial’ /gáy/ ‘tour guide’ /chumchon/ ‘community’ /sùanklaaṅ/ ‘centre’ /lookaaphíwát/ ‘globalisation’ /sǎmniaṅ/ ‘accent’ /nùayràachákaan/ ‘government office’ /pràpheenii/ ‘custom’ /sǎnlápàwátthánátham/ ‘art and culture’ /phûuyàybâan/ ‘village headman’ /hàaṅklay/ ‘far away’ /pòkkhrɔ̀ɔṅ/ ‘govern’ /sáppáyaakɔ̀on/ ‘resources’ /dâṅdɔ̀əm/ ‘traditional’

/sùanthóŋthìn/ ‘local’ /nákkaanmuan/ ‘politician’ /khonnay/ /nɔɔ/ (acronym for ‘o’clock’) /wátthánátham/ ‘culture’ /rádàp/ ‘level’ /ʔoŋkɔɔn/ ‘company’ /wannákam/ ‘literary work’ /booraansàthãan/ ‘ancient remains’ /phuumípâak/ ‘region’ /thêetsàbaan/ ‘municipality’ /kòtkràsuan/ ‘ministerial order’ /nãŋsũuphim/ ‘newspaper’ /kamnan/ ‘village headman’ /sùanrũam/ ‘contribution’ /sàphaa/ ‘council’ /kaanlũuktân/ ‘election’ /cèttánaarom/ ‘intent’ /sòŋsǎəm/ ‘support’ /phũunbãan/ ‘local’ /phôɔkháa/ ‘merchant’ /phũunmuan/ ‘native’ /thũathũn/ ‘thoroughly’ /withĩichiiwít/ ‘lifestyle’ /pràthêetchâat/ ‘nation’ /ʔèekkàlák/ ‘identity’ /chonnàbot/ ‘countryside’ /lũaktân/ ‘elect’ /ʔànúrák/ ‘conserve’ /phũunam/ ‘leader’ /sǎəmsãan/ ‘enhance’ /thóŋthĩi/ ‘district’ /sàmaachík/ ‘member’ /sàkkàyáphâap/ ‘potential’ /chũachâat/ ‘race’ /phrêelãay/ ‘widespread’ /khêmkhěen/ ‘strong’ /bɔɔrĩbot/ ‘context’ /wátthùdĩp/ ‘raw material’ /ʔaasãasàmàk/ ‘volunteer’ /kràcaay/ ‘spread’ /phípĩththáphan/ ‘museum’

The collocates of *district* ranked by Log Ratio score are as follows:

red-light tendring hoeveler enumeration dhas cherwell sedgefield non-metropolitan montgomeryshire richmondshire kinross renfrew braintree hambleton lake snowdonia outlying columbia registry teesdale pennines stroud basingstoke metropolitan shire hampshire council attorney polling peak lothian dungannon ryedale borough hospice **rural** designated wakefield nurse edinburgh councillor judge perth v. highlands glasgow urban parte derbyshire vale auditor cambridgeshire piccadilly **slum** county manhattan planner health cumbria eastern neighbouring prosecutor registrar nursing baltic devon fashionable teesside commissioner federal colchester tokyo surrounding dale dundee coastal authority predominantly Strathclyde gp referral valley **diabetic** angeles hospital hills aberdeen sterling mining northumberland municipal paisley provincial census **psychiatric** electoral suffolk provider superintendent audit working-class recreation ex boundary somerset belfast surveyor inhabitant residential darlington board planning officer merge oxfordshire cornwall governor adjacent **remote** yorkshire dean border east parish essex province bristol san clinic registration situate league residence branch durham jurisdiction administrative organise court region dublin southern city comprise

6.5.1.3 /pràwàtsàat/ vs. *history*

The collocates of /pràwàtsàat/ ranked by Log Ratio score are as follows:

historical /koŋlɔɔ/ ‘wheel’ /sámrrɔɔ/ ‘repeat’ /booraankhàdii/ ‘archeology’ /nákbooraankhàdii/ ‘archeologist’ /mánútsàyáchâat/ ‘mankind’ /sĩsàtchánaalay/ (city’s name) /phuumlãŋ/ ‘background’ /máhãabandit/ ‘postgraduate’ /phuumísàat/ ‘geography’ /wátthùniyom/

‘materialism’ /ʔùttháyaan/ ‘national park’ /bòoribòt/ ‘context’ /krèt/
 ‘anecdote’ /chiiwápràwàt/ ‘biography’ /ʔiŋ/ ‘base on’ /ʔàksóonsàat/ ‘arts’
 /yaawnaan/ ‘for a long time’ /sánkhomwíttháyaa/ ‘sociology’ /tùlaa/
 ‘October’ /sīnlápàwátthánátham/ ‘art and culture’ /maanútsàyáwíttháyaa/
 ‘anthropology’ /râakŋâw/ ‘root’ /booraansàthään/ ‘ancient remains’
 /phâakwíchaa/ ‘department’ /phoŋsääwádaan/ ‘historical annals’
 /wíwátthánaakaan/ ‘evolution’ /caarúk/ ‘inscribe’ /châatníyom/
 ‘nationalism’ /châatthay/ ‘Thailand’ /sīn/ ‘art’ /phíphíttháphan/ ‘museum’
 /chonchâat/ ‘nationality’ /hân/ ‘Hun Dynasty’ (part of a person’s name)
 /máak/ ‘Marxism’ /kâwkèε/ ‘ancient’ /phanpii/ ‘1,000 years’ /khónkhwáa/
 ‘study intensively’ /sámăykòon/ ‘in the past’ /bantúk/ ‘record’
 /râatcháwoŋ/ ‘dynasty’ /tamnaan/ ‘legend’ /lákthään/ ‘evidence’ /pràtyaa/
 ‘philosophy’ /caarít/ ‘custom’ /náváníyaay/ ‘novel’ /bòtrian/ ‘lesson’
 /sùkhòthay/ (city’s name) /yúk/ ‘era’ /ŋêemum/ ‘viewpoint’ /wíchaa/
 ‘subject’ /sīnlápà/ ‘art’ /bùkbàæk/ ‘pioneer’ /râaŋraaw/ ‘account’
 /wátthánátham/ ‘culture’ /châattíphan/ ‘ethnic’ /láanna/ (kingdom’s
 name) /sámăymà/ ‘modern time’ /thát/ ‘sight’ /níphon/ ‘writing’
 /sàthâapàtthàyákam/ ‘architecture’ /yóon/ ‘turn back’ /hèetkaan/ ‘event’
 /rátsia/ ‘Russia’ /ròŋròoy/ ‘trace’ /kriik/ ‘Greece’ /pàrinyaathoo/ ‘master’s
 degree’ /khunkhâa/ ‘value’ /mít/ ‘dimension’ /yīnyà/ ‘great’ /rátthàsàat/
 ‘political science’ /wannákádii/ ‘literature’ /râakthään/ ‘foundation’
 /chœŋ/ ‘base on’ /yúròop/ ‘Europe’ /nák/ (grammatical particle) /tamraa/
 ‘textbook’ /paarít/ ‘Paris’

The collocates of *history* ranked by Log Ratio score are as follows:

tartt 300-year **chequered** regius post-independence rewrit prehistory
 turning-point geography genealogy **revisionist** hardie ecclesiastical
 dialectical archeology tripos rewrite mankind potted recorded illustrated
 pelican unparalleled bede vol. dialectic watershed landmark anthropology
 topographical footnote delve annal milestone evolutionary philosophy
natural vertebrate mythology architectural oral chronological botany
 biography thermal aesthetics turning sociology economic **turbulent**
 snapshot literary epoch agrarian geological art museum chronology
 shipbuilding twentieth-century judaism reconstruct discipline ancient
 literature maritime dustbin concise fascinating civilisation astronomy
 bibliography episode social economics marx science trace reformation
 marxist aviation post-war medieval **slavery** **penal** brief humanities
 antiquity recount salvation legend modern distinguished steep
 psychoanalysis unfold colourful geology **ulcer** throughout lesson
 feminism culture developmental recent **madness** f. smoking subject
 biblical cultural sexuality **psychiatric** architecture lecturer narrative
 teaching unknown eighteenth-century mathematics anglo-saxon professor
diabetes theology english feminist nineteenth-century andrews airfield
 linguistics victoria myth making medical curriculum universe biology
glorious **tragic**

6.5.1.4 /thoorásàp/ vs. *telephone*

The collocates of /thoorásàp/ ranked by Log Ration score are as follows:

/thooráphim/ ‘teletype’ /**dàkfaŋ**/ ‘eavesdrop’ /meeloodii/ ‘melody’
 /thoorásään/ ‘fax’ /tít/ (ringing sound) /kríŋ/ (ringing sound) /thoorálêek/
 ‘telegraph’ /bəə/ ‘number’ /kríŋ/ ‘bell’ /khlúanthii/ ‘mobile’ /määylêek/
 ‘number’ /nämpràpaa/ ‘tap water’ /khlúanráp/ ‘receiver’ /bèet/ ‘battery’
 /khwían/ ‘throw’ /khâanám/ ‘water bill’ /kòt/ ‘dial’ /cháat/ ‘charge’ /sim/
 ‘sim card’ /**khàtcaŋwà**/ ‘interrupt’ /hũu/ ‘ear’ /sòophtháam/ ‘inquire’
 /krápáwthũu/ ‘handbag’ /pùm/ ‘button’ /pràpaa/ ‘water supply’ /pêen/
 ‘keypad’ (person’s name) /mee/ ‘email’ /nátmäay/ ‘arrange an
 appointment’ /thäämthây/ ‘ask about’ /foon/ ‘telephone’ /bèettəəríi/
 ‘battery’ /khuu/ ‘talk’ /han/ ‘hello’ /**khroom**/ ‘bang’ /sòŋkhàaw/ ‘send a
 message’ /kròok/ ‘fill in’ /**kriit**/ ‘shrill’ /sàmùt/ ‘notebook’
 /thooràsàpmuuthũu/ ‘mobile phone’ /sányaan/ ‘signal’ /thoo/ ‘telephone’
 /dâythii/ /**phèet**/ ‘shrill’ /khòotua/ ‘beg off’ /khrúanjkhòomphiwtəə/
 ‘computer’ /cǐw/ ‘tiny’ /tittòo/ ‘contact’ /khwáa/ ‘grab’ /khòoyuum/
 ‘borrow’ /săay/ ‘line’ /meen/ ‘email’ /khwaan/ ‘search for’ 0 /yip/ ‘pick’
 /nêep/ ‘enclose’ /tũu/ ‘box, cabinet’ /wítháyú/ ‘radio’ /praysánii/ ‘post’
 /**róŋriian**/ ‘complain’ /lúan/ ‘pick’ /còt/ ‘write down’ /mũn/ ‘dial’ /lòo/
 ‘hello’ /phúutkhuu/ ‘chat’ /còŋ/ ‘reserve’ /waan/ ‘put’ /?intənèt/ ‘internet’
 /tittân/ ‘install’ /**kràthêek**/ ‘bang’ /săatharána/ ‘public’ /còtmăay/ ‘letter’
 /khrúan/ ‘tool’ /?àtthànóomát/ ‘automatic’ /cêeŋ/ ‘notify’ /miisían/ ‘make
 sound’ /ríp/ ‘rush’ /dan/ ‘make sound, loud’ /hăa/ ‘see’ GAP:name_list
 /maaráyâat/ ‘manner’ /fayfáa/ ‘electricity’ /sían/ ‘sound’ /còo/ ‘screen’
 /**khũu**/ ‘threaten’ /sònthánaa/ ‘talk’

The collocates of *telephone* ranked by Log Ratio score are as follows:

direct-dial minibar toll-free **shrill** cordless hairdryer hotline dial ringing
 kiosk dial directory tea/coffee cellular taped bidder face-to-face helpline
 long-distance handset answering nippon facsimile daytime switchboard
 on-site telegraph mouthpiece postage conversation 24-hour **intercept**
 subscriber transatlantic doorbell ring stationery balcony portable fax
 modem at&t typewriter mobile caller **bug** cable radio bt tv digital receiver
 state-owned address postal wireless electricity tap call wire via bell
 bathroom install box anonymous transcript bedroom telecommunication
 telecom operator satellite number exchange booking booth desk co
 enquiry television jersey network selling notify mail **emergency** message
 installation line shower interview contact pad equip pick connect banking
warning facility corp **monopoly** letter telephone answer equipment
 heating toilet **urgent** electronic alarm bath communication hook

6.5.1.5 /phûukhon/ vs. *people*

The collocates of /phûukhon/ ranked by Log Ratio score are as follows:

/khlaakhlâm/ ‘go together in a group’ /**phlúkphlâan**/ ‘in disorder’ /**sôŋsũm**/ ‘assemble (in conspiracy)’ /mâaknâalâaytaa/ ‘too many people’ /baaŋtaa/ ‘few people’ /**khwàkkhwàw**/ ‘helter-skelter’ /**kwàattôn**/ ‘herd’ /**khraa**/ ‘take by force’ /**cœcœ**/ ‘crowded and noisy’ /rótraa/ ‘vehicle’ /**biatsiat**/ ‘congested’ /**nêenkhanàt**/ ‘packed’ /nâataa/ ‘many people’ /yûatyaan/ ‘vehicle’ /**têektuun**/ ‘panic’ /lànjlây/ ‘flow’ /**ŋonlâmàan**/ ‘confusing’ /**nuanênên**/ ‘crowded’ /phóppà/ ‘meet’ /sâncœn/ ‘travel’ /dænsũan/ ‘walk in the opposite direction’ /**khápkhâŋ**/ ‘crowded’ /hômmlóm/ ‘gather round’ /**ŋòtyàak**/ ‘starve’ /**khúkkhák**/ ‘vigorous’ /**ráaŋ**/ ‘deserted’ /**ŋùkkàtúk**/ ‘noisy’ /**yátyiat**/ ‘force’ /**ŋeŋàt**/ ‘congested’ /càpcàay/ ‘buy’ /bâanruan/ ‘house’ /cè/ ‘meet’ /withiichiiwít/ ‘lifestyle’ /**dũnduut**/ ‘attract’ /súuhãa/ ‘buy’ /**nãanênên**/ ‘crowded’ /**pliaw**/ ‘out-of-the-way’ /**muŋ**/ ‘crowd around’ /**lõŋlây**/ ‘infatuated’ /**ŋiãpŋãw**/ ‘lonely’ /mâakmaay/ ‘many’ /thaáyœy/ ‘gradually’ /**lòlían**/ ‘nourish’ /thâamklaaŋ/ ‘among’ /**kràcàt**/ ‘scattered’ /namphaa/ ‘bring’ /hèe/ ‘flock in, parade’ /**biat**/ ‘push through’ /**lóm**/ ‘die, fall’ /**kwâanyây**/ ‘vast’ /**yímyém**/ ‘cheerful’ /hàaŋklay/ ‘far away’ /**chiiwítchiiwaa**/ ‘lively’ /ŋòppháyóp/ ‘emigrate’ /**sàpsõn**/ ‘confused’ /làaklây/ ‘various’ /bâanmuãŋ/ ‘country’ /**thammaahãakin**/ ‘make a living’ /phonlãmuaŋ/ ‘citizen’ /phûanfũuŋ/ ‘friend’ /**wãaŋplàw**/ ‘empty’ /wêetlóm/ ‘surround’ /máhãasãan/ ‘huge’ /**nápthũu**/ ‘respect’ /**rum**/ ‘beset, crowd around’ /kràap/ ‘prostrate oneself, ship’s side’ /pràatsàcàak/ ‘without’ /**náamcay**/ ‘kindness’ /thûa/ ‘throughout’

The collocates of *people* ranked by Log Ratio score are as follows:

azanian azapo non-disabled colored right-thinking pre-literate ctsp hural kampuchean penan naacp manipur togolese **disabled** right-minded upp **colonised** roomful 4,000,000 eritrean elderly **handicapped** like-minded 10,000,000 rpt lao 900,000 **hiv/aids** npc **hiv-positive** mongolian **homeless** 1,500,000 mentally npa 110,000 2,000,000 **deaf** **disordered** **hospitalize** **oppressed** 42,000 indigenous able-bodied ppp **non-literate** **marginalised** well-educated **disability** 400,000 **killed** 90,000 **bereaved** **poverty-stricken** ÖVP 180,000 100,000 300,000 throng 250,000 50,000 150,000 **unemployed** commissariat 7,000 1,000,000 **massacre** 130,000 **displaced** sudanese **frail** 80,000 tribal 270 pensionable young 1,700 **enslave** millions liberation **injure** 700,000 well-off **chronically** 1,600 swapo siberian 500,000 600,000 **marginalise** 1,200 **low-income** well-intentioned 60,000 **infect** 18,000 30,000 **dementia** **behead** retired 350,000 vietnamese 200,000 **hiv** afro-caribbean ordinary thousands 4,500 5,000 khmer **overweight** 40,000 **disable** educated unicameral **infirm** 3,000 **kill** ethiopian **uneducated** fewer **starving** 8,000 10,000 20,000 trusting 25,000

emigrate hundreds aboriginal tens cameroon wanting self-determination
 employ 15,000 2,500 2,000 livelihood 4,000 **destitute needy** slav 1,500
 estimated 70,000 **obese invalidity** well-meaning anti-fascist **pester** 14,000
 12,000 **detain alienate** nomadic **disadvantaged impaired** 000 1,000
 employing knowledgeable self-employed exhort **persecute** many
starvation inhabit **grieving**

6.5.1.6 /phâapháyon/ vs. *film*

The collocates of /phâapháyon/ ranked by Log Ratio score are as follows:

film /klôŋthây/ ‘movie camera’ /ʔeɛnímeechân/ ‘animation’ /thâytham/
 ‘film’ /hɔɔnlí/ ‘Hollywood’ /thèɛpbanthúksǎŋ/ ‘magnetic sound-
 recording tape’ /namsáɛɛŋ/ ‘play the leading role’ /chǎay/ ‘show’
 /sǎarákádii/ ‘documentary’ /wúut/ ‘Hollywood’ /phûukamkàp/ ‘director’
 /rooŋ/ ‘hall’ /wùut/ ‘Hollywood’ /wiidiiʔoo/ ‘video’ /phâapnîŋ/ ‘slide’
 /fim/ ‘film’ /wiidii/ ‘DVD, CVD, video tape’ /chom/ ‘watch’ /daaraa/
 ‘movie star’ /phûuchom/ ‘audience’ /cɔɔ/ ‘screen’ /thooráthát/ ‘television’
 /núarûaŋ/ ‘content’ /thèetsàkaan/ ‘festival’ /khôotsànaa/ ‘advertisement’
 /banthəəŋ/ ‘entertainment’ /sàmaa/ ‘federation’ /yɔɔtyíam/ ‘excellent’
 /náksáɛɛŋ/ ‘actor’ /dòoŋdan/ ‘famous’ /líkee/ (Thai theoretical
 performance) /phan/ ‘federation’ /sǐŋphim/ ‘printed matter’ /sùu/ ‘media’
 /təlòk/ ‘funny’ : /wíttháyú/ ‘radio’ /waarásǎan/ ‘journal’ /núahǎa/ ‘content’
 /woŋkaan/ ‘field’ /ʔùtsǎahàkam/ ‘industry’ /dàtplɛɛŋ/ ‘adapt’ /hɔɔ/
 ‘building’ /bûaŋlǎŋ/ ‘behind the scenes’ /lákhoon/ ‘drama’ /rûaŋsǎn/ ‘short
 story’ /thát/ ‘sight’ /níttháyásǎan/ ‘magazine’ /naanaachâat/ ‘international’
 /pràchaasǎmphan/ ‘public relations’ /chûudan/ ‘famous’ /wícaan/
 ‘criticise’ /múan/ ‘roll’ /roomeɛntik/ ‘romantic’ /chàak/ ‘scene’ /cəə/
 ‘meet’ /bòt/ ‘screenplay, role’ /thàlɛɛŋkhàaw/ ‘press conference’
 /phráʔèek/ ‘leading actor’ /sǎaŋ/ ‘produce’ /khàawsǎan/ ‘news’ 35
 /kaatuun/ ‘cartoon’ /théknik/ ‘technique’ /dontrii/ ‘music’ /phûuphàlit/
 ‘producer’ /pràphêet/ ‘type’ /chàlǎəm/ (part of a place name) /thay/ ‘Thai’
 /khrua/ ‘network’ /náwáníyaay/ ‘novel’ /kawlii/ ‘Korea’ /naaŋʔèek/
 ‘leading actress’ /théep/ ‘tape’ /camnàay/ ‘sell’ /phǎəyphrɛɛ/ ‘disseminate’
 /bɔɔrísàt/ ‘company’ /thúrakit/ ‘business’ the /sàyǎam/ ‘Siam’
 /nǎŋsǔuphim/ ‘newspaper’

The collocates of *film* ranked by Log Ratio score are as follows:

antonioni 16mm billson cine **low-budget** puttnam eisenstein spielberg
 35mm newsreel cannes scorsese greenaway fuji tie-in oscar-winning
porno footage amadeus **pornographic** soundtrack **noir** kodak
 schwarzenegger **gangster** hitchcock chaplin remake premiere bergman
 spoof mgm box-office celluloid projector rambo acetate polyester

hollywood documentary olivier walt branagh starring attenborough emulsion **porn** screenplay subtitle brando cinema festival beatty **ensor** black-and-white python jurassic showing award-winning eastwood disney buff film-maker co-star woody clint cowboy cling polyethylene ealing reviewer video animated cameraman blockbuster maker producer script television star gung transistor photographic archive animation realist oily comedy bbc2 **horror** photography **thriller** tv screening fu mokul close-up shanghai distributor camera crew sequel making reel **stunt** audrey transparent studio critic columbia actress warner broadway actor lesbian x-ray madonna viewing film **flop** preview clip dub cartoon abdominal oscar depict movie thin moving steven cassette mainstream narrative **ensorship** silent theatre feature director epic **propaganda** fiction gay academy projection capture newman classic screen **violent** shoot industry jungle vietnam cult début drama portray sean soap sequence slide viewer legend resulting british protective polymer broadcast watch hero poster adventure classification production strip edit loop allen roll opera 1930s

6.5.1.7 /pràsithíphâap/ vs. *efficiency*

The collocates of /pràsithíphâap/ ranked by Log Ratio score are as follows:

efficiency /pràsittiphôn/ ‘success’ /pràsit/ ‘achievement’ /khlôŋtua/ ‘flexible’ /thûathǎn/ ‘thoroughly’ /pròŋsǎy/ ‘transparent’ /phêmphuun/ ‘improve’ /thanthûanthii/ ‘in time’ /dôy/ ‘inferior’ /sǎomsâan/ ‘enhance’ /pràyàt/ ‘thrifty’ /sàmátthàná/ ‘capacity’ /phêem/ ‘increase’ /bòorihǎan/ ‘administer’ /ʔèekkàphâap/ ‘unity’ /chonpràthaan/ ‘irrigation’ /sùusǎan/ ‘communicate’ /thótsòp/ ‘test’ /sáppháyaakoon/ ‘resources’ /sũunsùt/ ‘highest’ /pràpprun/ ‘improve’ /phàlit/ ‘produce’ /sámsóon/ ‘overlapping’ /càtsǎn/ ‘allocate’ /damnəŋŋaan/ ‘operate’ /ráy/ ‘without’ /phálanŋaan/ ‘energy’ /kaanbòorikaan/ ‘service’ /phûupàtibàtŋaan/ ‘officer’ /khúmkhâ/ ‘worthwhile’ /rûatrew/ ‘quick’ /fukʔòprom/ ‘train’ /pàtibàtŋaan/ ‘work’ /sàmátthàphâap/ ‘capability’ /kàsèettàkoon/ ‘agriculturist’ /thansàməy/ ‘modern’ /cháyŋaan/ ‘use’ /sǎankheemii/ ‘chemical substance’ /pràməŋ/ ‘estimate’ /kamcat/ ‘get rid of’ /cháycaay/ ‘spend’ /bòorihǎan/ ‘administer’ /phianphoo/ ‘adequate’ /sòŋsǎem/ ‘support’ /khrûaŋcàk/ ‘machine’ /konkay/ ‘mechanism’ /nǎanmaacàak/ ‘due to’ /càtkaan/ ‘manage’ /thamŋaan/ ‘work’ /sàthianráphâap/ ‘stability’ /plòtphay/ ‘safe’ /khàat/ ‘lack’

The collocates of *efficiency* ranked by Log Ratio score are as follows:

allocative cost-effectiveness effectiveness aerodynamic improving optimise renewable energy **impair** productivity improve improved maximize competitiveness computational boiler profitability equity

rationality increased operational thermal managerial brisk optimum labelling improvement accountability scrutiny productive morale saving **ruthless** reliability administrative appliance fuel promote fairness incentive economy operating enhance flexibility maximum pursuit gain **decrease** accuracy increase allocation boost increasing conversion economic indicator evaluate conversation technical competence measure nhs emphasis assess achieve speed agricultural manufacturing **reduce** safety taxation economics internal overall monitor great output peak reduction greatly competition consideration funding enterprise cost imply quality electricity industrial

6.5.1.8 /*η̄ankhăy*/ vs. *condition*

The collocates of /*η̄ankhăy*/ ranked by Log Ratio score are as follows:

condition /lækkeən/ ‘principle’ /*η̄anweelaa*/ ‘time clause’ /*phòonpron*/ ‘ease’ /*dunláyáphâap*/ ‘balance’ /*kòtkràsuaŋ*/ ‘ministerial order’ /*bòoríbòt*/ ‘context’ /*wêetlóm*/ ‘surround’ /*khòokamnòt*/ ‘specification’ /*phaaytây*/ ‘under’ /*pràkùatraakhaa*/ ‘bid’ /*sămrèt*/ ‘succeed’ /*wíthiikaan*/ ‘method’ /*sòmrot*/ ‘marry’ /*fàafüun*/ ‘violate’ /*khòocamkàt*/ ‘limitation’ /*kamnòt*/ ‘specify’ /*tòopayníi*/ ‘from now on’ /*pràphrút*/ ‘behave’ /*pràatsàcàak*/ ‘without’ /*lonthòot*/ ‘punish’ /*ʔàthípòodii*/ ‘director-general’ /*nítikam*/ ‘juristic act’ /*sàmákaan*/ ‘equation’ /*sàphâapkaan*/ ‘state’ /*pràmuun*/ ‘bid’ /*khum*/ ‘control’ /*ʔûaʔamnuay*/ ‘facilitate’ /*tòmûa*/ ‘when’ /*koo*/ (letter /*koo* kày/) /*khêmŋûat*/ ‘strict’ /*phrátikaan*/ ‘behaviour’ /*kòtkeen*/ ‘regulation’ /*pàtibàt*/ ‘perform’ /*khonthii*/ ‘unchanged’ /*pràkaan*/ ‘point’ /*tòoròŋ*/ ‘negotiate’ /*ʔànúyâat*/ ‘allow’ /*câaŋ*/ ‘employ’ /*sămpàthaaŋ*/ ‘concession’

The collocates of *condition* ranked by Log Ratio score are as follows:

tip-top **insanitary** **handicapping** **parlous** **life-threatening** **unhygienic** causally **disabling** **atrocious** **deteriorating** immaculate climatic mint treatable **inhumane** **deteriorate** prevailing pristine **blustery** **waterlogged** **worsening** meteorological **cramped** **overcrowded** **worsen** **recessionary** **choppy** anaerobic **appalling** favourable **ruinous** **unfavourable** **psychosomatic** labouring implied stringent sanitary **windy** **adverse** **overcrowd** ameliorate **deplorable** **incurable** bail optimum **foggy** pre-existing **osteoporosis** conducive **inhuman** neurological alkaline humid **arthritic** weather aetiology precedent **eczema** **onerous** humidity **aggravate** enabling fulfil **pathological** acidic satisfied visibility **squalid** **arid** satisfy **unsafe** **stressful** **epilepsy** warranty **impose** prevail **blizzard** ambient changed stipulate **imposition** optimal **distressing** working stabilise **unhealthy** excellent simulated socio-economic equilibrium

breach boundary **drought** concurrent **treacherous** **chronic** **arduous**
comply experimental physiological living improved **harsh** altered
arthritis atmospheric controlled **freezing** **spartan** under qualifying
inflammatory improve recreate **deterioration** attach respiratory **wartime**
arctic **wretched** aquarium moist spawn diagnose physical **diabetes** **poor**
improving neutral **restrictive** changing eligibility **unstable** adapt
abnormal **crowded** trading stable entry employment viewing **damp**
simulate tolerate **oppressive** rectify **exacerbate** specify **imposing**
workhouse **strict** perfect ideal ecological alleviate **intolerable** covenant
unsuitable **deviation** specified medical **extreme** **asthma** geological
rigorous satisfactory thrive engagement govern **severe** sufficient
workplace **primitive** ph **icy** **critical** **psychiatric** economic certain
environmental threshold soft necessary term wet initial

6.5.1.9 /phûuyà/ vs. *adult*

The collocates of /phûuyà/ ranked by Log Ratio score are as follows:

/phûunóoy/ ‘junior’ /sùukhǎo/ ‘propose marriage’ /kəəntua/ ‘beyond one’s
power’ /rápwây/ ‘return good will’ /chûafan/ ‘obey’ /naaytháhãan/
‘soldier’ /khûnnaañ/ ‘nobleman’ /temtau/ ‘completely’ /yâat/ ‘relative’
/ʔəənnóom/ ‘submissive’ /ʔammàat/ ‘court official’ /khawróp/ ‘respect’
/ʔùppàthãm/ ‘patronise’ /nôəpnóom/ ‘show deep respect’ /ʔuayphóon/
‘bless’ /rótnám/ ‘pour water on’ /ʔenduu/ ‘have compassion for’
/sěenaabóodii/ ‘minister’ /khûukhróon/ ‘spouse’ /way/ ‘age’ /wayrûn/
‘teenage’ /khaaráwá/ ‘respect’ /təəptoo/ ‘grow’ /waan̄tua/ ‘behave’
/naaytamrúat/ ‘policeman’ /kamnan/ ‘village headman’ /mēettaa/
‘goodwill’ /chomchəəy/ ‘praise’ /dèk/ ‘child’ /ʔaawúsóo/ ‘senior’
/sànsǎon/ ‘teach’ /caydii/ ‘kind’ /nápthūu/ ‘respect’ /phûuban̄khápban̄chaa/
‘chief’ /câwnaay/ ‘boss’ /khâarâatchákaan/ ‘government official’ /thamtua/
‘act’ /khâwhãa/ ‘approach’ /câwkun/ (personal pronoun) /chán/ ‘level’
/too/ ‘grow up’ /lianbèep/ ‘copy’ /phôm̄mêe/ ‘parents’ /tón̄nãa/ ‘in the
presence of’ /khèek/ ‘guest’ /sũun̄ʔaayú/ ‘elderly’ /wây/ ‘pay respect’
/tóontôn/ ‘first part’ /sóm/ ‘equivalence’ /taamcay/ ‘spoil’ /phûuthâw/
‘aged man’ /kreen̄cay/ ‘considerate’ /lâk/ ‘main’

The collocates of *adult* ranked by Log Ratio score are as follows:

consenting suffrage £1.5 £2.5 literacy **immature** £3.5 juvenile universal
50p education **handicapped** larva five-year learner adolescent £6 £2
adolescence £4 dependant mature £3 plumage continuing **mortality** alike
mentality £1 £5 beetle homosexual male childhood emerging child healthy
unemployed survivor middle-aged **offender** sexually infant tutor young
disabled competent miniature population working-class butterfly

disability franchise offspring teenager **infection** admission female elect household learning 1,000 supervision breeding physically life advisory aged accompany fare provision skilled **disorder** volunteer ratio part-time **acute** sexual behave

6.5.1.10 /sùkkhàphâap/ vs. *health*

The collocates of /sùkkhàphâap/ ranked by Log Ratio score are as follows:

/phálaanaamay/ ‘good health’ /sǔw sǔw sǔw/ (acronym for ‘Thai Health Promotion Foundation’) /ʔànaamay/ ‘hygiene’ /sâaŋsǔəm/ ‘enhance’ /khǎeŋrɛeŋ/ ‘strong’ /lákpràkan/ ‘security’ /sàmátchaa/ ‘assembly’ /sùkkhàphâapcìt/ ‘mental health’ /sútsoom/ ‘worn out’ health /bànthǔwŋ/ ‘weaken’ /ʔǔkkamlan̄kaay/ ‘exercise’ /rúk/ ‘approach’ /phòotcháanaakaan/ ‘nutrition’ /thûan/ ‘all’ /ráwan/ ‘careful’ /phuumphéε/ ‘allergy’ /duuleε/ ‘take care of’ /kaankiilaa/ ‘sports’ /kaay/ ‘body’ /cèpkây/ ‘ill’ /sàycay/ ‘attentive’ /sǔəmsâan/ ‘enhance’ /cèppùay/ ‘ill’ /pràkan/ ‘guarantee’ /sàwàt/ ‘safety’ /màn/ ‘often’ /sùamsoom/ ‘declining’ /kaanbǔwrikaan/ ‘service’ /fit/ ‘fit’ /truat/ ‘check’ /ʔantàraay/ ‘danger’ /phàlittàphan/ ‘product’ /sàpaa/ ‘spa’ /ràan̄kaay/ ‘body’ /ʔǔkkamlan̄/ ‘exercise’ /khrûaŋdùum/ ‘beverage’ /phīwphan/ ‘complexion’ /sǔŋsǔəm/ ‘support’ /hùan̄yay/ ‘care about’ /sìŋwêetlǔwŋ/ ‘environment’ /sǔmbuun/ ‘perfect’ /sǎathaarànásùk/ ‘public health’ /thīyùuʔaasǎy/ ‘accommodation’ /fūunfuu/ ‘restore’ /phǔnsǎa/ ‘damage’ /bambàt/ ‘cure’ /mâysûu/ ‘not quite’ /phǔwǔm/ ‘appropriate’ /phǔnkràthóp/ ‘effect’ /sàmàttàphâap/ ‘capability’ /sàmŋphray/ ‘herb’ /ràksǎa/ ‘cure’ /phákphǔwŋ/ ‘rest’ /kǔwŋthun/ ‘fund’ /sūuŋʔaayú/ ‘elderly’ /khúkkhaam/ ‘threaten’ /phūubǔwriphòok/ ‘consumer’ /plǔwŋphay/ ‘safe’ /pàtirûup/ ‘reform’ /khúmkrǔwŋ/ ‘protect’ /thaarók/ ‘baby’ /sàtipanyaa/ ‘wisdom’ /dii/ ‘good’ /sǔŋphǔn/ ‘affect’ /rǔok/ ‘disease’ /châat/ ‘nation’ /damrǔŋchīiwít/ ‘live a life’

The collocates of *health* ranked by Log Ratio score are as follows:

pallot timmins gawor dhas bottomley dept. coshh mental blunkett dorrell mawhinney fhsa midwifery jerzy safety tees medicare waldegrave wellbeing nih sanitation yeo hydro **injurious** occupational preventive cohse preventative virginia nutrition check-up **hazard** midwife authority fitness care ssds nupe **deteriorating ill** two-tier service watchdog wessex **detrimental** promotion welfare **inequality** harman holistic visitor hygiene maternal wholeness environmental dental unicef clinic mersey nutritional **endanger** 1936 immunisation antenatal **deteriorate** shake-up regional epidemiology district **morbidity** **hazardous** reimburse lothian education gp improving hse department dhss under-secretary educator nurse

multidisciplinary nursing spokeswoman rationing ministry nhs clwyd ayrshire powys correspondent smoking vaccination spa grampian medicine 1875 **ill-health chronic** national insurance baroness **failing** executive **sickness** greater **impair** confederation **warning** kenneth vitality practitioner mohammed secretary physician **deterioration nuisance** privatise public directorate improved **overweight declining** purchasing maternity staffordshire lifestyle primary ancillary determinant board provider **cholera** workplace prevention social community oxfordshire **harmful risk** dalington screening worker happiness **surveillance impairment** behavioural promote fletcher officer **scare aids** physical reproductive medical visiting professional dentist **mortality** institute equitable maximise spokesman improve inspectorate socio-economic berkshire indicator **illness** clarke essex provision restore campaigner organization guideline 1959 **epidemic disability** junior thames 1946 **acute** beauty beneficial referral economics regulation

6.5.1.11 /rátthàthammánuun/ vs. *constitution*

The collocates of /rátthàthammánuun/ ranked by Log Ratio score are as follows:

221 291 /tùlaakaan/ ‘judge’ 266 /sǒǒ sǒǒ rǒǒ/ (acronym for ‘Member of Drafting Constitution Council’) /ràatcháʔaanaacàk/ ‘kingdom’ 190 178 /bòtchápókaan/ ‘transitory provision’ 67 /pǒǒrámaaphíthay/ ‘king’s name’ /sǎan/ ‘court’ /ràaŋ/ ‘draft’ /phúthásàkkàràat/ ‘Buddhist era’ 2540 /sàmaachíkkàphâap/ ‘membership’ /cèettànaarom/ ‘intent’ 50 /sǎanpòkkhrǒǒŋ/ ‘The Administrative Court’ /wínitchǎy/ ‘judge’ 2550 /pràchaamátì/ ‘referendum’ /kêekhǎy/ ‘amend’ /sǎanyúttitham/ ‘Court of Justice’ /lómláaŋ/ ‘overthrow’ /banyàt/ ‘enact’ /bòtbanyàt/ ‘provision’ /khàt/ ‘go against’ /phràràatchábanyàt/ ‘act’ /chàbàp/ ‘edition’ /kaanlûaktân/ ‘election’ /yúp/ ‘dissolve’ /rátthàsàphaa/ ‘parliament’ /khánákammaathíkaan/ ‘commission’ 2475 /prákǒǒp/ ‘supplement’ /phǒǒ rǒǒ bǒǒ/ (acronym for ‘act’) 66 /yéeyŋ/ ‘oppose’ /kammaathíkaan/ ‘commissioner’ /pràmúk/ ‘leader’ /wúthhísàphaa/ ‘Senate’ 48 /ràbǒǒp/ ‘system’ /phákkaanmuay/ ‘political party’ /yáttì/ ‘motion’ /chúikhàat/ ‘make a final decision’ /phràràatcháthaa/ ‘give’ /mâattraa/ ‘section’ /leekhánúkaan/ ‘secretary’ /sàphaaphûtheenràatsàdǒǒn/ ‘House of Representatives’ /sàphaa/ ‘council’ /pràthaa/ ‘chairman’ /phràmaháakàsàt/ ‘great king’ /khǒǒ sǒǒ/ (acronym for ‘Christian era’) /kòtʔayyákaansùk/ ‘martial law’ /wák/ ‘paragraph’ /pǒǒ pǒǒ chǒǒ/ (acronym for ‘Office of the National Anti-Corruption Commission’) /sàmátchaa/ ‘assembly’ /laaylák/ ‘letter’ /phaaytây/ ‘under’ /lámôǒt/ ‘violate’ /sǎeriiphâap/ ‘freedom’ /sǒmbuuránaayaasithíráat/ ‘absolute monarchy’ /pàtirûup/ ‘reform’ /càttham/ ‘make’ /sàmaachík/ ‘member’ /kòtmǎay/ ‘law’ /pràthaaanaathípǒǒdii/ ‘president’ /phôǎmtǎǎm/ ‘increase’

/náptèè/ ‘since’ /hě́nchôp/ ‘approve’ /fàafűun/ ‘violate’ /pràkàat/ ‘announce’ /kàtikaa/ ‘rule’ /pràchaathíppàtay/ ‘democracy’ /hě̀n/ ‘of’ /sítthimánútsàyáchon/ ‘human right’ /nítibanyat/ ‘legislation’ /yók/ ‘raise’ 50 /khôôsàñǎə/ ‘proposal’ /yóklǎək/ ‘cancel’ /conkwàa/ ‘until’ /ʔitsàrà/ ‘freedom’

The collocates of *constitution* ranked by Log Ratio score are as follows:

unwritten promulgate promulgation enshrine amend draft preamble amended multiparty referendum stipulate **violate** **contravene** drafting amendment written 1937 **suspend** enact federal **prohibit** adoption namibia transitional 1947 approve revision interim 1917 revise republic **forbid** embody ussr supersede vest ratify adopt **violation** 1918 genetic cork czech guarantee 1962 legislative republican presidential revised accordance proposed constitutional 1960 clause kingdom monarchy 1936 democratic decree endorse legislature irish **abolish** liberty 1980 article approval sovereignty proclaim new 1978 statute **suspension** 1976 envisage 1966 alter supreme liberal united british 1977 states assembly formally devise **criticize** supervision provision soviet convention 1972 nov. law electoral president congress under executive parliament pacific parliamentary 1974 1973 contrary freedom election fundamental recognize russian 1982 **restrict** restore reform

6.5.1.12 /pâwmăay/ vs. *target*

The collocates of /pâwmăay/ ranked by Log Ratio are as follows:

goal goals target /banlú/ ‘achieve’ /chíiwát/ ‘indicate’ /wátthùpràsǒŋ/ ‘purpose’ /ŋə̀nfǎə/ ‘inflation’ /sǎmrít/ ‘achieve’ /yútthàsàat/ ‘strategy’ /phǒnsǎmrèt/ ‘achievement’ /phûukǎkaanráay/ ‘terrorist’ 1.3 /nêechát/ ‘clearly’ /còcoŋ/ ‘specify’ /sũũsùt/ ‘highest’ /phě̀nŋaan/ ‘project’ /phûurian/ ‘learner’ /klùm/ ‘group’ /khâwthǎŋ/ ‘access’ /pràchaakǎoŋ/ ‘people’ /thiitǎŋ/ ‘location’ /sǎ̀otkhlǎoŋ/ ‘harmonious’ /damnə̀nŋaan/ ‘operate’ /phûufan/ ‘audience’ /mûŋ/ ‘intend’ /khâatwǎŋ/ ‘hope’ /khwaamsǎmrèt/ ‘success’ /chátceen/ ‘clear’ /waan/ ‘set’ /mêenyam/ ‘accurate’ /coomtii/ ‘attack’ /keen/ ‘criterion’ /konláyút/ ‘strategy’ /plaaythaan/ ‘destination’ /tòkpen/ ‘become’ /phǒnláp/ ‘result’ /sǎmphâat/ ‘interview’ /náyoobaay/ ‘policy’

The collocates of *target* ranked by Log Ratio score are as follows:

attainment stabilising 0.7 attainable 2005 achievable take-over intended \$10 moving **vandal** emission ambitious recycle stabilize exceed shareholder £250,000 realistic profitability acquisition cbi attain tempting

recycling stringent legitimate dioxide **terrorist** opec 1996 agreed header takeover **unrealistic demanding** flare monetary civilian hot align **bomb** 2000 ira audience sitting achieve carbon sulphur chosen **bombing** £10,000 vendor antigen reduction set specific recruitment manpower spending imf **thief** indicator shooting revise setting ie meet specified prime reach quota missile selected modest **attack** acquire iraqi dna profile achievement accuracy identification **raid** share obvious short-term completion sales identity bang intervention easy israeli performance asset **deficit**

6.5.1.13 /chaaynùm/ vs. *young man*

The collocates of /chaaynùm/ ranked by Log Ratio score are as follows:

/kamyam/ ‘strong’ /sǎnthát/ ‘medium-sized’ /ʔâmʔûŋ/ ‘equivocate’ /lâmsǎn/ ‘sturdily built’ /khlǝkhlia/ ‘caress’ /khomkhaay/ ‘good-looking’ /tiinâa/ ‘pretend’ /krâaw/ ‘rough’ /sǎøy/ ‘brush back’ /kraam/ ‘molar’ /sàbòt/ ‘swear’ /thúm/ ‘deep sound’ /pháyákpháyâët/ ‘give a nod as a signal’ /ʔùkʔàk/ ‘stammering’ /cǝŋmǝŋ/ ‘stare’ /chǝt/ ‘shirt’ /praay/ ‘glance’ /hûak/ ‘heavily’ /sêŋ/ ‘pretend’ /chamluaŋ/ ‘glance’ /phlǝŋ/ ‘blurt out’ /buan/ ‘turn’ /tǝktàluŋ/ ‘dumbfounded’ /sǎŋʔǝŋ/ ‘soft voice’ /tàwàt/ ‘swing’ /pháwaŋ/ ‘subconsciousness’ /lǝp/ ‘sneakily’ /chúachǝn/ ‘invite’ /phǝm/ ‘thin’ /phumpham/ ‘mumble’ /khrǝm/ ‘reserved’ /khraaŋ/ ‘moan’ /tǝtbòt/ ‘break off a discussion’ /yím/ ‘smile’ /thamʔaw/ ‘cause’ /yǎn/ ‘jeer’ /ʔomyím/ ‘smile in mildly manner’ /rápkham/ ‘promise’ /phêŋ/ ‘stare’ /yókmǝu/ ‘raise one’s hand’ /khrûnkhit/ ‘brood’ /pháyák/ ‘nod’ /râaŋ/ ‘figure’ /prǝŋ/ ‘slender’ /cǝpcǝŋ/ ‘stare’ /lamkhǝw/ ‘neck’ /hûan/ ‘abruptly’ /ràapriap/ ‘smooth’ /lùap/ ‘glance’ /lanlee/ ‘hesitate’ /ʔùtcay/ ‘one moment’ /làptaa/ ‘close one’s eyes’ /lǝmmtaa/ ‘open one’s eyes’ /hǎnlǎŋ/ ‘turn one’s back towards’ /sǎinâa/ ‘facial expression’ /lǝwlian/ ‘mimic’ /ʔûŋ/ ‘dumbfounded’ /thǝŋ/ ‘withdraw’ /rii/ ‘narrow’ /lǎw/ ‘sharpen’ /khiw/ ‘eyebrow’ /phlaaŋ/ ‘at the same time’ /khlám/ ‘tanned’ /naytaa/ ‘eye’ /yǝ/ ‘jeer’ /nâataa/ ‘face’ /ŋǝy/ ‘look up’ /cháják/ ‘stop abruptly’ /sǎay/ ‘shake’ /khàmùat/ ‘knot’ /hǎn/ ‘turn’ /liaw/ ‘glance’ /khonkhap/ ‘driver’ /sút/ ‘sink’ /wǝwtaa/ ‘eyes expression’ /khǝwtua/ ‘beg off’ /phiŋ/ ‘lean on’ /klân/ ‘hold back’ /sǝdûŋ/ ‘startled’ /nítnuŋ/ ‘a little’ /ʔèey/ ‘utter’ /sùut/ ‘snuff’ /kum/ ‘hold’ /yûunmǝu/ ‘stretch out one’s hand’ /kôm/ ‘bend down’ /phǝŋ/ ‘loosen’ /baynâa/ ‘face’ /námsǎŋ/ ‘tone’ /sǝp/ ‘meet’ /ʔùthaan/ ‘exclaim’ /khàyp/ ‘move’ /duanŋtaa/ ‘eye’ /nîŋ/ ‘silent’ /khêm/ ‘dark, deep’ /kràsíp/ ‘whisper’ /khûumân/ ‘fiancé’ /plèeknâa/ ‘strange’ /hûarǝ/ ‘laugh’ /thûinân/ ‘seat’ /ʔǝnyoon/ ‘gentle’ /ʔendu/ ‘have compassion for’ /rǝyyím/ ‘smile’ /plèekcay/ ‘surprised’ /phùt/ ‘rise, occur’ /tǝpráp/ ‘answer, accept’ /khom/ ‘attractive’ (part of a person’s name) /lày/ ‘shoulder’ /sǎamsíp/ ‘thirty’ /yǝm/ ‘stand’ /rimphǝpàak/ ‘lips’ /yák/ ‘shrug’ /nûmnuan/ ‘gentle’ /mǝŋ/

‘look at’ /khrûu/ ‘moment’ /lomhăaycay/ ‘breath’ /wûup/ ‘momentary’ /prăkhəŋ/ ‘carry’ /phlăk/ ‘push’ /kreençay/ ‘considerate’ /baw/ ‘softly’

The collocates of *young man* ranked by Log Ratio score are as follows:

personable good-looking dark-haired handsome blond twenties earnest ambitious **angry** tall charming intelligent beard ambition wealthy eager pleasant smart brave shirt clever healthy bright portrait nice **pale** dress

6.5.1.14 /sɨŋwêetlóm/ vs. *environment*

The collocates of /sɨŋwêetlóm/ ranked by Log Ratio score are as follows:

/thəw sɔw/ (acronym for ‘Ministry of Natural Resources and Environment’) /sáppháyaakəŋthammáchâat/ ‘natural resources’ environmental environment /ʔànúrăk/ ‘conserve’ /**phǒnkràthóp**/ ‘effect’ /**sùamsoom**/ ‘declining’ /**khunnápâap**/ ‘quality’ /ʔánaamay/ ‘hygiene’ /**mít**/ ‘friendly’ /**monphaawá**/ ‘pollution’ /phuumíʔaakàat/ ‘climate’ /kammáphan/ ‘heredity’ 2535 /níwêet/ ‘residence’ /**monláphít**/ ‘pollution’ /**yânyuun**/ ‘sustainable’ /phanthúkam/ ‘heredity’ /**bamrunráksăa**/ ‘maintain’ /wítsawákam/ ‘engineering’ /níwêetwítháyaa/ ‘ecology’ /**sùkkhăaphíbaan**/ ‘sanitation’ /krəsuaŋ/ ‘ministry’ /**sòŋsəəm**/ ‘support’ /**sǒmdun**/ ‘balance’ /**pûan**/ ‘dirty’ /cătkaan/ ‘manage’ /**phíthák**/ ‘protect’ /kaayyáphâap/ ‘physical’ /phëen/ ‘plan’ /wítháyaaasàat/ ‘science’ /pătisămphān/ ‘interaction’ /**phayphibăt**/ ‘disaster’ /**khǒŋsăa**/ ‘waste’ /**thamlaay**/ ‘destroy’ /téknoołooyii/ ‘technology’ /rátthàmontriiwâakaan/ ‘minister’ /sàhàpràchaachâat/ ‘United Nation’ /**răksăa**/ ‘cure’ /**sùkkhàphâap**/ ‘health’ /**sínlápawátthánátham**/ ‘art and culture’ /**khúmkhəŋ**/ ‘protect’ /châat/ ‘nation’ /kòohâykəət/ ‘cause’ /khánákammákaan/ ‘committee’ /**hěnychòp**/ ‘approve’ /pon/ ‘mix’ /**tòkkháaŋ**/ ‘left over’ /ʔoŋpràkòp/ ‘composition’ /**prăptua**/ ‘adjust’ /thíiyùʔaasăy/ ‘accommodation’ /**sáppháyaakəŋ**/ ‘resources’ /**muunníthí**/ ‘foundation’ /pràməŋphǒn/ ‘evaluate’ /**kràthóp**/ ‘affect’ /**cítsămnúk**/ ‘conscious’ /pălăt/ ‘deputy’ /thammáchâat/ ‘nature’ /**mâattràthăan**/ ‘standard’ /**phít**/ ‘poison’ /khánákammaathíkaan/ ‘commission’ /**săathaarănásùk**/ ‘public health’ /**phálanŋaan**/ ‘energy’ /**kəŋthun**/ ‘fund’ /sămnákŋaan/ ‘office’ /khamnəŋ/ ‘consider’ /raayŋaan/ ‘report’ /**plòtphay**/ ‘safe’ /wikhró/ ‘analyse’ /naanaachâat/ ‘international’ /tălòtcon/ ‘including’ /**răk**/ ‘protect’ /**ʔantàraay**/ ‘danger’ /phûut/ ‘plant’ /phûubòríphòok/ ‘consumer’ /phaaynòk/ ‘outside’ /**ronnărəŋ**/ ‘campaign’

The collocates of *environment* ranked by Log Ratio score are as follows:

Ségolène auslan cramb Töpfer lutzenberger lalonde ripa gef trippier baldry unep brice distributed client-server softbench bouchard nextstep clustered

unced character-based client/server dme dce computing ryley object-oriented c++ nerc windows-based osf **ever-changing** klaus **pollute** graphical sna heseltine cobol tivoli heredity taligent run-time embedded carlo x-windows conducive yeo emil engineered doe unix-based hp-ux gummer workgroup stimulating humid toolkit freshwater director-general uniform terrestrial patten department heterogeneous interconnect **hostile** maclean networked multimedia netware man-made real-time tcp/ip ridley salim east-west california-based **polluted unfavourable** solaris operating conserve interact programming adapt os/2 aix desktop organism supportive macintosh arid stockholm disciplined distribute unix homely networking clean-up changed safe hp **unhealthy** changing physical **watery** marine integrated commissioner global competitive protect correspondent controlled 3.0 **harmful degradation** ms-dos harriet **unfamiliar** software enhancement ecology aquatic mainframe **turbulent** urban natural proprietary un tourism howard circular hewlett-packard adaptation fragile transportation working microsystem simulate 2.0 **contamination** learning deposition external **sterile** stable regulatory **harsh** foundation integrate bryan portable challenging protection novell **contaminate** caring **hazard** favourable motif interactive secretary nt processing reporter **manipulate** healthy mac **deterioration** **adverse** sustainable pleasant **noisy** minister co-operative surrounding **unstable** thrive application ecological development safeguard ec **rural** warming wildlife coastal dynamic hardware workstation interface structured maintain **alien** unveil visual **pollution** microsoft digital **uncertain** gould database michael conservation enhance advanced agency intel impact create ministry commons classroom immediate inc organizational environmental improve agriculture climate shadow specification chris di technological beneficial corp developer

6.5.1.15 /pràsòpkaan/ vs. *experience*

The collocates of /pràsòpkaan/ ranked by Log Ratio score are as follows:

experience /sànsòm/ ‘accumulate’ /kèpkìaw/ ‘gather’ /sànsǎəm/ ‘enhance’ /chíawchaan/ ‘expert’ /phêəmphuun/ ‘improve’ /rôəprúu/ ‘well-rounded’ /khwaamrúu/ ‘knowledge’ /bèənpaŋ/ ‘share’ /phuumlǎŋ/ ‘background’ /lêəkplian/ ‘exchange’ /rianrúu/ ‘learn’ /chamnaaŋ/ ‘skilled’ /khlúkklíi/ ‘associate intimately with’ /thàaythôət/ ‘pass on’ /bòtrian/ ‘lesson’ /tháksà/ ‘skill’ /khǒmkhùun/ ‘painful’ /lòokkàthát/ ‘one’s view of the world’ /chæ/ ‘share’ /sǎəmsàaŋ/ ‘enhance’ /wíchaachîp/ ‘vocation’ /leewráay/ ‘horrible’ /khunnáwút/ ‘qualification’ /sàsòm/ ‘accumulate’ /fúkfǒn/ ‘practise’ /wíttháyaakəən/ ‘lecturer’ /phûupàtibàtŋaan/ ‘officer’ /phuumpanyaa/ ‘intelligence’ /yaawnaaŋ/ ‘for a long time’ /cintànaakaan/ ‘imagination’ /cèppuàt/ ‘painful’ /pràthápcaj/ ‘impressed’ /sámphàt/

‘feel’ /khwaamsăamâat/ ‘ability’ /thâap/ ‘compare’ /dôoy/ ‘inferior’ /khrañnán/ ‘at that time’ /bùkkálíkkàphâap/ ‘personality’ /sùantua/ ‘personal’

The collocates of *experience* ranked by Log Ratio score are as follows:

near-death firsthand post-qualification hands-on cathartic **harrowing** **traumatic** humbling **nerve-wracking** **hair-raising** **hallucinatory** on-the-job **sobering** mystical vicarious perceptual pleasurable **unsettling** accumulated unforgettable **unnerving** relive rewarding subjective shared sensory **shattering** exhilarating visionary uplifting invaluable formative salutary seniority **terrifying** **unpleasant** jimi **searing** **frightening** everyday enjoyable unrivalled expertise recount liberating memorable gain whitehouse sharing **humiliating** parenting **painful** breadth learning fishkeeping practical childbirth pre-school aesthetic sensual mathematical hendrix qualification real-life **distressing** valuable mystic placement **stressful** previous **frustrating** broaden glean childhood profoundly know-how stimulating totality **horrific** knowledge psychic **depressing** bodily satisfying personal enrich transcend **wartime** theatrical **daunting** **bitter** **bereavement** wisdom moving uniquely emotional **disturbing** learner intuition clinical ageing varied wealth learn educational visual perception religious diversity undergo spiritual imaginative sexual extensive skill lifetime relevant innocence differing collective operational unique articulate competence **lack** derive flying profound past

6.5.1.16 /khwaamtôñkaan/ vs. *desire*

The collocates of /khwaamtôñkaan/ ranked by Log Ratio score are as follows:

needs need /tòopsànđõñ/ ‘fulfil’ /sànđõñ/ ‘fulfil’ /sànđõñtòop/ ‘fulfil’ /săanchâattàyaan/ ‘instinct’ /sđõtkhlđõñ/ ‘harmonious’ /phianphõc/ ‘adequate’ /khânp húunthăan/ ‘basic’ /phunphõccay/ ‘satisfied’ /təəmtem/ ‘fulfil’ /ʔùppàphôok/ ‘consume’ /rõõñráp/ ‘support’ /khântàm/ ‘minimum’ /tanhăa/ ‘lust’ /phûubõoríphôok/ ‘consumer’ /fayfăa/ ‘electricity’ /bõoríphôok/ ‘consume’ /tàlàat/ ‘market’ /thátsánákáti/ ‘attitude’ /rótníyom/ ‘taste’ /námmandìp/ ‘crude oil’ /còot/ ‘question’ /phêet/ ‘gender’ /pàrímaan/ ‘quantity’ /sínchûa/ ‘credit’ /cuunçay/ ‘persuade’ /phûucháy/ ‘user’ /phõophian/ ‘sufficient’ /lûukkháa/ ‘customer’ /kâawráaw/ ‘aggressive’ /phûunthăan/ ‘foundation’ /yùurđõc/ ‘survive’ /pàtcèek/ ‘individual’ /cháysđõy/ ‘use’ /rëeñjaan/ ‘labourer’ /pràchaachon/ ‘people’ /sũunçùt/ ‘highest’ /ʔõõm/ ‘save (money)’ (person’s name) /pálanñaan/ ‘energy’ /khwaansõncay/ ‘interest’ /làaklăay/ ‘various’ /naykhànán/ ‘at that time’

The collocates of *desire* ranked by Log Ratio score are as follows:

streetcar **insatiable unfulfilled unsatisfied** motivate **repress** longing **overwhelming** sensual fulfilment **perverse ardent insane** emulate **earthly** satisfy **obsessive thwart earnest** innate sexual **burning frustrated** irresistible stifle override **lust revenge** passionate impulse curiosity instinctive **unconscious** outweigh understandable fulfil express fantasy **restless urgent jealousy** happiness genuine consciously **shudder selfish** arouse surge **fierce** minimise **hunger** pulse inherent intense **suppress** passion object affection consume reflect mutual satisfaction ambition reinforce communicate satisfied unit unity prompt motive **provoke** flame pleasure belief strong **desperate** sudden stem emotion heart **dominate** possess strengthen flesh escape

6.5.1.17 /tàwantòk/ vs. *west*

The collocates of /tàwantòk/ ranked by Log Ratio score are as follows:

/chǎŋtây/ ‘southward’ western /thit/ ‘direction’
 /sittthisàphâapnôkʔaanaakèet/ ‘extraterritorial rights’ /chǎŋnǎa/
 ‘northward’ /khaawbɔɔy/ ‘cowboy’ west /chǎŋ/ ‘diagonal’ /tawanʔôk/
 ‘east’ /ʔaaráyátham/ ‘civilisation’ /máhǎaʔamnâat/ ‘great power’ /sǐik/
 ‘side’ /sǎalá/ (part of a river’s name) /càkkàwàtníyom/ ‘imperialism’
 /chaayfǎŋ/ ‘coast’ /thátthiam/ ‘equal’ /fǎŋ/ ‘side’ /thúakkhǎw/ ‘mountain
 range’ /paakiisàthǎan/ ‘Pakistan’ /yúròop/ ‘Europe’ /mɔɔrásǔm/
 ‘monsoon’ /lǔóklian/ ‘copy’ /pɛesífik/ ‘Pacific’ /làŋlǎy/ ‘flow’
 /ʔaanaanikhom/ ‘colony’ /ʔaaráyá/ ‘civilised’ /phèekhàyyǎay/ ‘expand’
 /ʔaawthay/ ‘the Gulf of Thailand’ /nákkhít/ ‘thinker’ /ʔéepfikaa/ ‘Africa’
 /sàmmǎymà/ ‘modern time’ /muanǎkhûn/ ‘colony’ /ʔandaaman/
 ‘Andaman’ /bèepyàan/ ‘model’ /nákpàat/ ‘learned man’ /pràtyaa/
 ‘philosophy’ /khòon/ ‘incline’ /wátthánátham/ ‘culture’ /protùkèet/
 ‘Portugal’ /khòop/ ‘edge’ /phàsǔmphasǎan/ ‘integrate’ /lamnáam/ ‘river’
 /ʔittthíphon/ ‘influence’ /rooman/ ‘Roman’ /khúkkhaam/ ‘threaten’ /fǎak/
 ‘side’ /wíthiikhít/ ‘mindset’ /thèep/ ‘zone’ /chaaw/ ‘citizen’ /wíphǎak/
 ‘comment’ /thansǎmǎy/ ‘modern’ /kiloóméet/ ‘kilometer’ /khwɛɛ/ (river
 and city’s name) /hǎnnâa/ ‘face’ /monthon/ ‘county’ /nákwíchaakaan/
 ‘scholar’ /ʔóottreelia/ ‘Australia’ /càròt/ ‘stretch to’ /máhǎasàmùt/ ‘ocean’
 /yǎeráman/ ‘Germany’ /kɔɔ mɔɔ/ (acronym for ‘kilometre’) /khròppǎam/
 ‘overpower’ /phêet/ ‘doctor’ /kriik/ ‘Greece’ /tháwǐip/ ‘continent’ /mùukò/
 ‘group of islands’ /dindɛɛn/ ‘territory’ /rúkraan/ ‘invade’ /ràtsia/ ‘Russia’
 /ʔeechia/ ‘Asia’ /ʔirǎan/ ‘Iran’ /wannákhádii/ ‘literature’ /wayyaakɔɔn/
 ‘grammar’ /ʔindia/ ‘India’ /nǎa/ ‘north’ /látthí/ ‘doctrine’ /sàttàwát/
 ‘century’ /thunníyom/ ‘capitalism’ /phuumpanyaa/ ‘intelligence’ /dontrii/
 ‘music’ /lianbèep/ ‘copy’ /sàhàrátʔameeríkkaa/ ‘America’ /lòok/ ‘world’

/thálee/ ‘sea’ /tàwan/ ‘sun’ /fáa/ ‘sky’ /thaaŋ/ ‘side’ /pràchaathíppàtay/ ‘democracy’ /mêenáam/ ‘river’ /wíwátthánaakaan/ ‘evolution’ /chaaydeen/ ‘frontier’

The collocates of *west* ranked by Log Ratio score are as follows:

indies bromwich brom torridge penwith sussex-based ham gaza ecowas
 mae midlands-based midlands israeli-occupied penistone york mercia
 berliner bengal gillick riding witton marland wisbech nat shipley
 littlehampton albion germany hans-dietrich glamorgan kirby belfast
 stourbridge ardiles idian ossie coast monklands draytom cumbria lothian
 croydon sowerby saxon helmut thurrock harrow sussex lewisham
 hampstead beirut yorkshire bundesbank hartlepool highland german
 genscher berlin lancs somerset grinstead brentfort acton brompton
 keighley anglia kensington rhine brandt virginia solihull midland oban
 east north crawley falkirk richie fulham hammersmith sumatra Paddington
 africa kingston clwyd abingdon wakefield viv kohl chiswick pennine
 ealing auckland south m4 cumberland african falkland derby bundestag
 middlesex mercian embassy java rebecca dudley farther hebrides
 playhouse bonn bowler herts angling norfolk dorset km millwall settler
 prague barnsley hants greenland swansea south-west lancashire
 humberside siemens shetland suffolk mountains reunification shropshire
 surrey clayton huddlesfield coalfield **occupied** timothy spd durham mp
 Bristol bournemouth bradford netherlands siberia palestinian asian
 bowling chichester willy berkshire jerusalem wales innings bank thames
 wolverhampton **wild** devon leicester halifax strip vs belgium exodus
 london hughes cornwall highlands scotland syrian malaysia cheshire
 peninsula chancellor jamaica warsaw regent one-day bloc north-west loch
 severn migrant unification mains **con** austria frontier atlantic merseyside
 italy guinea richardson ridge heath refuge coventry mile france derbyshire
 sutton cliff wicket denmark plantation hampshire south-east israeli kent
 japan exeter sweden end richards wiltshire richmond greater islands
 canada

6.5.1.18 /ʔitthíphon/ vs. *influence*

The collocates of /ʔitthíphon/ ranked by Log Ratio score are as follows:

influence /sàkàtkân/ ‘intercept’ /phèekhàyyäy/ ‘expand’ /câwphôo/
 ‘godfather’ /khrôppam/ ‘overpower’ /bìpbaŋkháp/ ‘force’
 /cittâysămnúk/ ‘subconscious’ /phêe/ ‘expand’ /sàatsánacàk/ ‘religious
 ruling authority’ /soowîat/ ‘Soviet’ /rátthàsàat/ ‘political science’
 /lookaaphíwát/ ‘globalisation’ /mɔ̀rásũm/ ‘monsoon’ /rooman/ ‘Roman’
 /khǒm/ ‘Khmer’ /thátsánákháti/ ‘attitude’ /tawantòk/ ‘west’ /hinduu/

‘Hindu’ /sə̀muanchon/ ‘mass media’ /tāanthaan/ ‘withstand’ /tòɔ/ ‘to’ /látthít/ ‘doctrine’ /pàtcay/ ‘factor’ /sàphâapwêetlóm/ ‘surroundings’ /máhãaʔamnâat/ ‘great power’ /khàcàt/ ‘get rid of’ /wíthiikhít/ ‘mindset’ /lòokkàthát/ ‘one’s view of the world’ /phûuson/ ‘holder’ /nǎa/ ‘over’ /khàmëen/ ‘Khmer’ /sǎjkhomwítháyaa/ ‘sociology’ /cìtsǎmnúk/ ‘conscious’ /khâaniyom/ ‘value’ /kriik/ ‘Greece’ /ʔaaràp/ ‘Arab’ /khə̀mmiwnít/ ‘communism’ /phútthásàatsànãa/ ‘Buddhism’ /khwaamkhít/ ‘thought’ /ʔaaráyátham/ ‘civilisation’ /dâyrap/ ‘receive’ /nákkhít/ ‘think’ /phaaytây/ ‘under’ /citwítháyaa/ ‘psychology’ /nákkhít/ ‘thinker’ /wátthánátham/ ‘culture’ /phaayú/ ‘storm’ /phaaynòk/ ‘outside’ /khàyyáay/ ‘expand’ /pràtyaa/ ‘philosophy’ /dooytroŋ/ ‘directly’ /khwaamchûa/ ‘belief’ /sàatsànãa/ ‘religion’ /wannákhàdii/ ‘literature’ /sòŋ/ ‘demand’ /phrúttikam/ ‘behaviour’ /kaanmuang/ ‘politics’ /ʔindia/ ‘India’ /tāaŋchâat/ ‘foreign’ /sə̀usāanmuanchon/ ‘mass communication’ /bùkkhalíkkàphâap/ ‘personality’ /wêetlóm/ ‘surround’ /ʔùppà/ ‘demand’ /kanlékan/ ‘to each other’ /sàpeen/ ‘Spain’

The collocates of *influence* ranked by Log Ratio score are as follows:

moderating **undue** stabilizing civilizing exert **corrupting** steadying restraining formative calming **pernicious** pervasive moorish wield **malign** sphere byzantine germanic **disruptive** planetary profound **peddle** determining predominant decisive **wane** shaping counteract **disproportionate** discernible lasting controlling potent maternal superpower **destructive** amenable powerful parental exercise dominant climatic considerable benign overriding guiding **declining** diverse indirect **weaken** **diminish** external causal persian maximise continental cultural upon genetic immense growing regain strong **communist** strengthen **overwhelming** immune beneficial under enormous significant extend continuing jewish jazz bargaining musical alcohol ideological perception wealth political **undermine** magnetic intellectual christianity greek undoubtedly direct **negative** cite combined italian

6.5.1.19 /chûamooŋ/ vs. *hour*

The collocates of /chûamooŋ/ ranked by Log Ratio score are as follows:

/kiloowàt/ ‘kilowatt’ 24 4-6 /thamweelaa/ ‘take time’ /may/ ‘mile’ /kinweelaa/ ‘take time’ /khrâŋ/ ‘half’ 48 /ʔoosoon/ ‘ozone’ /kilooméet/ ‘kilometre’ /yíisip/ ‘twenty’ 72 /pràkhóp/ ‘massage’ 1-2 /naathii/ ‘minute’ 120 /kii/ ‘how many’ 72 2-3 3-4 /kùap/ ‘almost’ /pràmaan/ ‘approximately’ /raaywan/ ‘daily’ /sii/ ‘four’ /sèet/ ‘fraction’ /pèet/ ‘eight’ /chàliã/ ‘average’ /khwaamrew/ ‘speed’ 12 45 /raaw/ ‘approximately’ 12 /khâap/ ‘period’ /wínaathii/ ‘second’ /lá/ ‘each’ /kɔɔ mɔɔ/ (acronym for

‘kilometre’) /sàpdaa/ ‘week’ /hòŋrian/ ‘classroom’ /nè/ (pragmatic particle) /weelaa/ ‘time’ /chêe/ ‘soak’ /yàaŋnóoy/ ‘at least’ /khâachàlia/ ‘mean’ /naan/ ‘for a long time’ /khâacâaŋ/ ‘wage’ /kəən/ ‘more than’ /yaawnaaŋ/ ‘for a long time’ 40 /khôŋ/ ‘more than half’ /rêŋdùan/ ‘urgent’ /tòcàaknán/ ‘then’ /bin/ ‘fly’ /lăŋcàaknán/ ‘after that’ /khrûaŋbin/ ‘plane’ /wanyùt/ ‘holiday’ /hòk/ ‘six’ /tûuyen/ ‘refrigerator’ /tâlòot/ ‘throughout’ 70 /khêm/ ‘hand (of clock)’ /phaaynay/ ‘within’ /sàŋâambin/ ‘airport’ 8 /náptèe/ ‘since’ /sìp/ ‘ten’ /təŋcháv/ ‘morning’ 1 /náp/ ‘count’ /săam/ ‘three’

The collocates of *hour* ranked by Log Ratio score are as follows:

24-48 **unsocial** witching 0600 £3.40 twenty-four forty-eight 0700 permitted 1¾ 1¼ **unsociable** kilowatt 48 0900 one-and-a-half **ungodly** waking 24 marinate thirty-six two-and-a-half **lateness** 365 2½ three-and-a-half 37°C 3½ **unearthly** 4½ eleventh daylight 1½ 72 1100 seventy-two 10.00 chime kip half sunrise 0800 helpline 1/2 36 kick-off 1600 simmer weekday infusion 2-3 three-quarters spend incubation incubate non-stop 240 night-time ingestion twilight tuition 1-2 appointed chargeable overtime twenty-three later 3-4 **gruelling** scheduled **anti-social toil endurance** thirty-five ½ working intravenous mile 120 £1.50 daytime **fruitless** approximately equalise solidly gallon opening flying couple **rush** kilometre injection 160 twelve browse beforehand 11th 12 sunset noon sleep deliberation p.m. quarter awake midnight lunch dusk few forty-five within ph visiting a.m. licensing eight polling **onset tedious** morning after 140 dawn minute sunshine bake **ordeal** flexible extended 180 clock four spare several duration £3 **weary** overnight darkness 45 nearly six 2.5 queue an 35 marathon chill countless **snatch irregular** 200

6.5.1.20 /phúunthăan/ vs. *foundation*

The collocates of /phúunthăan/ ranked by Log Ratio score are as follows:

fundamental basic /săathaarânuupâphôok/ ‘public utility’ /nâmmannlòolêun/ ‘lubricating oil’ /nêewnáyooobaay/ ‘policy’ /khròŋsâaŋ/ ‘structure’ /ʔaaráyátham/ ‘civilisation’ /ŋəŋfəə/ ‘inflation’ /lâkkaan/ ‘principle’ /prátyaa/ ‘philosophy’ /müaŋrêe/ ‘mine’ /phuumlăŋ/ ‘background’ /thrítsà/ ‘theory’ /khánítsàat/ ‘maths’ /pàtcay/ ‘factor’ /phəəphiŋ/ ‘sufficient’ /bon/ ‘on’ /puu/ ‘lay’ /càthăa/ ‘provide’ /sâŋchâattàyaan/ ‘instinct’ /khwaamrúu/ ‘knowledge’ /thrítsâdii/ ‘theory’ /tháksà/ ‘skill’ /damrəŋchiiwít/ ‘live a life’ /phuumpanyaa/ ‘intelligence’ /khwaamchăa/ ‘belief’ /khunnátham/ ‘virtue’ /nêewkhít/ ‘idea’ /chonprâthaan/ ‘irrigation’ /ʔəŋpràkòəp/ ‘composition’ /săarásönthêet/ ‘information’ /khönsòŋ/ ‘transport’ /wíchaa/ ‘subject’ /khámánaakhom/

‘communication’ /khântàm/ ‘minimum’ /rɔŋráp/ ‘support’ /khâaniyom/
 ‘value’ /khwaamkhít/ ‘thought’ /khwaamtɔŋkaan/ ‘desire’ /sàthiti/
 ‘statistics’ /sámāanchǎn/ ‘united’ /bûaŋtôn/ ‘initial’ /kaanlonthun/
 ‘investment’ /dāŋdǎəm/ ‘traditional’ /sèetthàkìt/ ‘economy’ /thâwthiam/
 ‘equal’ /bèèpcamlɔŋ/ ‘model’ /cìtwìtháyaa/ ‘psychology’
 /khwaampencin/ ‘reality’ /cìtsǎmnák/ ‘conscious’ /sǎŋkhomniyom/
 ‘socialism’ /ʔanpen/ /wìthiikìt/ ‘mindset’ /kaayyáphâap/ ‘physical’
 /plùukfǎŋ/ ‘inculcate’

The collocates of *foundation* ranked by Log Ratio score are as follows:

gulbenkian sef menil wolfson nuffield wellcome mellon guggenheim
 laying dme non-profit carnegie charitable rowntree epistemological dce
 nsf lay software monastic **shaky** computing tivoli nutrition osf/1 distribute
 strait distributed **insecure** macroeconomic masonry solid solomon motif
 barnes powder moore concrete stone lasting graphical open theoretical
 conceptual mechanic hewlett-packard **undermine** trustee anniversary
scholarship prosperity ulster environment science build medieval dig
 certificate charity 1948 **disabled** founder ideological sport core joseph
 mathematics religious rational sponsor cathedral art faith provided status
 unix upon henry educational brick establish pile

6.5.1.21 /sàthǎanákāan/ vs. *situation*

The collocates of /sàthǎanákāan/ ranked by Log Ratio score are as follows:

situation /khápkhǎn/ ‘critical’ /chùkchǎn/ ‘urgent’ /sùkɲəm/ ‘critical’
 /mìnmèe/ ‘precarious’ /khliikhlaay/ ‘ease up’ /ʔùthókàphay/ ‘flood’
 /lɔlǎəm/ ‘dangerous’ /náamthûam/ ‘flood’ 2548 /rápmuu/ ‘cope with’
 /leewráay/ ‘horrible’ /tuŋkhríat/ ‘tense’ /kɔ̀ɔ̀pkûu/ ‘save’ /ʔùbàttiphay/
 ‘accident’ /chàpónâa/ ‘urgent’ /sàŋkhlɔŋ/ ‘unstable’ /triamphrɔ̀ɔ̀m/ ‘get
 ready’ /camlɔŋ/ ‘simulate’ /pràməən/ ‘estimate’ /sûuróp/ ‘fight’
 /thanthûŋthii/ ‘in time’ /wǎyphríp/ ‘quick-witted’ /bìipbaŋkháp/ ‘force’
 /wêetlɔ̀ɔ̀m/ ‘surround’ /bâanmuəŋ/ ‘country’ /monlápht/ ‘pollution’
 /tittaam/ ‘follow’ /thâwthan/ ‘astute’ /tawanʔɔ̀ɔ̀kklaəŋ/ ‘Middle East’
 /wikrit/ ‘crisis’ /khwaamrunrɛəŋ/ ‘violence’ /sɔ̀ɔ̀tkhlɔ̀ɔ̀ŋ/ ‘harmonious’
 /pàri/ ‘context’ /phàchəən/ ‘encounter’ /naykhànàni/ ‘at this moment’
 /phâapruam/ ‘overall’ /khànàni/ ‘at this moment’ /léəŋ/ ‘drought’
 /pràptua/ ‘adjust’ /bɔ̀ɔ̀rìbòt/ ‘context’ /chênnii/ ‘like this’ /ʔhâʔamnuay/
 ‘facilitate’ /yâaklambàak/ ‘troublesome’ /khâywàt/ ‘flu’ /mòsòm/
 ‘suitable’ /tɔ̀ɔ̀psánɔ̀ɔ̀ŋ/ ‘fulfil’ /khûnyùukàp/ ‘depend on’ /sàphâapkaan/
 ‘state’ /phlík/ ‘change’ /khwaamkhlûanwǎy/ ‘movement’ /kòtdan/
 ‘pressure’ /klâychít/ ‘closely’ /rêŋdùan/ ‘urgent’ /sìthimánútsàyáchon/
 ‘human right’ /than/ ‘catch up with’ /kêekhǎy/ ‘improve’ /naykhànànan/

‘at that time’ /wítòk/ ‘worried’ /khâatkaan/ ‘predict’ /phǒnkràthóp/ ‘effect’ /sàṅòp/ ‘peaceful’ /chaaydæen/ ‘frontier’ /kòòkaan/ /kaanmuaŋ/ ‘politics’ /phâaktây/ ‘southern part’ /pàtcùban/ ‘at present’ /phâaytây/ ‘under’ /ráayræŋ/ ‘grave’ /khûapkhum/ ‘control’ /phàchənnâa/ ‘encounter’

The collocates of *situation* ranked by Log Ratio score are as follows:

no-win catch-22 defuse **worsening deteriorating worsen** real-life **stressful oversimplify intolerable ameliorate rectify misread destabilise exacerbate aggravate misjudge perilous deteriorate life-threatening deplorable farcical normalize threatening** changed **risky dire precarious hypothetical volatile tricky** one-to-one stabilize analogous reassess **unstable hopeless seriousness unsatisfactory fictitious complicate** recurring conducive communicative **complicated remedy clear-cut** simulated **absurdity succinctly** appraise **chaotic radically urgency** face-to-face **ludicrous embarrassing alleviate anomalous** arise **unforeseen kashmir desperate potentially clarify** given open-ended **escalate fraught** objectively Zaïre somalia afghanistan utterance **awkward bizarre** foresee improve tense abnormal redress alarming gravity dangerous cope confront everyday **confusing tragic problematic** regarding resolve dilemma dramatically favourable confused deterioration socio-economic reverse simplify handle hazardous analyse exposed whereby **explosive manipulate complexity** react realistic assess factual envisage conflict unfamiliar current comedy learning irony classroom day-to-day similar **hostage** fundamentally monitor **monopoly** transform prevail unhappy difficult uncertainty

6.5.1.22 /khamtòp/ vs. *answer*

The collocates of /khamtòp/ ranked by Log Ratio score are as follows:

/khâatkhán/ ‘force’ /nâapen/ /caycòtcaycò/ ‘attentively’ /chàlǎoy/ ‘answer’ /khónhãa/ ‘search for’ /khamthãam/ ‘question’ /ròokhòoy/ ‘wait’ /bèepsòpthãam/ ‘questionnaire’ /phlòŋ/ ‘blurt out’ /pristànáa/ ‘puzzle’ /thamʔaw/ ‘cause’ /nêechát/ ‘clearly’ /khòsòp/ ‘test’ /trâytròŋ/ ‘think over’ /sàay/ ‘shake’ /tòŋhãa/ /còot/ ‘question’ /sãmrèrúup/ ‘ready-made’ /sàwɛɛŋhãa/ ‘seek’ /thuŋ/ ‘ask for’ /sákthãam/ ‘inquire’ /khòsòŋsây/ ‘doubt’ /kràcàaŋ/ ‘clear’ /khrûnkhít/ ‘brood’ /rò/ ‘wait’ /pháyák/ ‘nod’ /kòkhũ/ /còcòŋ/ ‘specify’ /thɛen/ ‘replace’ /hãa/ ‘find out’ /tòŋkaan/ ‘want’ /khò/ (letter /khò/ khwaay/) /laŋlee/ ‘hesitate’

The collocates of *answer* ranked by Log Ratio score are as follows:

coldwater yes/no monosyllabic affirmative **glib conundrum truthful**
 postcard **evasive** question definitive unequivocal resounding satisfactory
 unspoken unambiguous clear-cut ready-made riddle correct emphatically
 elicit written emphatic partial quiz query **evade** suggested definite prayer
incorrect respondent simple convincing coupon summons straightforward
 obvious parliamentary oral questionnaire marine tropical **dilemma** honest
lie coherent acceptable

6.5.1.23 /raayláʔiat/ vs. *detail*

The collocates of /raayláʔiat/ ranked by Log Ratio score are as follows:

/pliikyôy/ ‘miscellaneous’ /sòpthääm/ ‘inquire’ /rûuppâphan/
 ‘description’ /phêmtæm/ ‘increase’ /dâythîi/ /cèekcæŋ/ ‘explain
 thoroughly’ /kròok/ ‘fill in’ /plæŋ/ ‘plan’ /khunnáláksàná/ ‘characteristic’
 /tátthôn/ ‘cut down’ http /sǎŋkhèep/ ‘briefly’ /thoo/ ‘telephone’
 /khrópthûan/ ‘completely’ /càttham/ ‘make’ /phěenthîi/ ‘map’ /còtcam/
 ‘remember’ /nêep/ ‘enclose’ /cæŋ/ ‘explain’ /chíicæŋ/ ‘explain’ 5.1
 /phànùak/ ‘append’ /sâycaŋ/ ‘attentive’ /pháunphîw/ ‘surface’ /tòpayníi/
 ‘from now on’ /kiawkàp/ ‘about’ /bantúk/ ‘note’ 2.1 /camnêek/ ‘classify’
 /pròot/ ‘please’ /chàlàak/ ‘label’ /daŋníi/ ‘as follows’ /phròomthán/
 ‘together with’ /hũakhôv/ ‘topic’ /klàawthǎŋ/ ‘mention’ /núahãa/ ‘content’
 /khântôn/ ‘procedure’ /còt/ ‘write’ /ʔèekkàsáan/ ‘document’ /wâacâaŋ/
 ‘employ’ /sákthääm/ ‘inquire’ /khôokamnòt/ ‘specification’ : /khón/
 ‘search’ /khwaamthùuktôn/ ‘justice’ /cò/ ‘delve’ /woŋŋæŋ/ ‘sum’
 /kârúnaa/ ‘mercy’ /phâapruam/ ‘overall’ /lúk/ ‘deep’ /taaraaŋ/ ‘table’
 /tòktèeŋ/ ‘decorate’ /pàetphǎøy/ ‘reveal’

The collocates of *detail* ranked by Log Ratio score are as follows:

0293 stockist tel meticulous **gory** constructional biographical 081 sae
sketchy finalise 8pm **sordid** 071 divulge overleaf **lurid painstaking**
gruesome scrupulous anatomical **unimportant** **coy** disclose update
 contextual examine close-up **omit** precise module leaflet itinerary factual
intimate glean discuss booking brochure full chapter postcard architectural
intricate enclose appendix confidential graphic vivid exact **trivial** attention
horrific contact far circulate describe page suffice excursion following
 procedural reservation supporting listing decorative reveal specify booklet
authentic analyse technical forthcoming concerning explore layout 1992

6.5.1.24 /phômêe/ vs. *parents*

The collocates of /phômêe/ ranked by Log Ratio score are as follows:

/pùuyâa/ ‘grandparents’ /yàakan/ /buntham/ ‘adoptive’ /yàaráaŋ/ ‘divorce’ /phîinóŋ/ ‘siblings’ /**kamphráa**/ ‘orphan’ /phrómnâa/ ‘all present’ /**chûufan**/ ‘obey’ /líanduu/ ‘look after’ /**thôttthíŋ**/ ‘abandon’ /phrákhun/ ‘benevolence’ /kátanyuu/ ‘grateful’ /pronníbàt/ ‘look after’ /**raŋkæe**/ ‘bully’ /sàŋsǒŋ/ ‘teach’ /hǔaʔòk/ ‘feelings’ /phûupòkkhrǒŋ/ ‘guardian’ /**taamcay**/ ‘spoil’ /kôtaamcay/ /**tháló**/ ‘quarrel’ /lûuk/ ‘child’ /câwsǎaw/ ‘bride’ /yâat/ ‘relative’ /**bunkhun**/ ‘debt of gratitude’ /**bâannôk**/ ‘rural’ /lûuklǎan/ ‘descendant’ /ʔawcaysà/ ‘pay attention’ /**kâawráaw**/ ‘aggressive’ /yâa/ ‘divorce’ /**lianbêep**/ ‘copy’ /bâankǎət/ ‘hometown’ /hǎaŋəən/ ‘make money’ /phîilían/ ‘nanny’ /phûanfûuŋ/ ‘friend’ /khâatwǎŋ/ ‘hope’ /dèk/ ‘child’ /phûuyà/ ‘adult’ /ʔòprom/ ‘instruct’ /ʔòpʔùn/ ‘genial’ /yaay/ ‘grandmother’ /khrǒpkhrua/ ‘family’ /thîirák/ /ʔûm/ ‘carry’ /ʔawcay/ ‘try to please’ /**khêmŋûat**/ ‘strict’ /phanthúkam/ ‘heredity’ /bùat/ ‘enter the monkhood’ /plùukfǎŋ/ ‘inculcate’ /**kreençay**/ ‘considerate’

The collocates of *parents* ranked by Log Ratio score are as follows:

adoptive lone doting step-parent grandparent foster sibling **divorced** **bereaved** child-rearing **heartbroken** **grieving** offspring parenting **divorce** adopted **disapprove** **murdered** surrogate bride prospective denise reunite **distraught** governor **break-up** **dissatisfied** unmarried child ‘would-be jennifer **handicapped** toddler parental teacher **disappoint** pupil teenager emigrate spouse self-help charter adolescent middle-class elderly **bland** birth relinquish ageing guardian loving cot **deaf** schooling working-class consent newsletter responsive ballot **anxious** separation **worried** inform **unhappy** **absent** carer financially whose wealthy actively reassure alike

6.5.1.25 /nɛwkhít/ vs. *idea*

The collocates of /nɛwkhít/ ranked by Log Ratio score are as follows:

/máaksít/ ‘Marxism’ Marx /mánútsà/ ‘human’ /sit/ ‘Marxism’ /sëeriiníyom/ ‘liberalism’ /**wátthùníyom**/ ‘materialism’ /máak/ ‘Marxism’ /phûurírêəm/ ‘originator’ /thrítsà/ ‘theory’ /nákthítsàdii/ ‘theorist’ /niiʔoo/ ‘neo’ /thrítsàdii/ ‘theory’ /dii/ ‘theory’ /**châatníyom**/ ‘nationalism’ /sǎŋkhomwítháyaa/ ‘sociology’ /phaasásàat/ ‘linguistics’ /pràtyaa/ ‘philosophy’ /kiawyoŋ/ ‘relate’ community /máhăayaan/ ‘Mahayana’ /sǎŋkhomníyom/ ‘socialism’ /nákpràtyaa/ ‘philosopher’ /kròp/ ‘framework’ /pàtcèek/ ‘individual’ /sǒmmúttithǎan/ ‘assumption’

/naammátham/ ‘abstract’ /frɔɔy/ ‘Freud’ /pràyúk/ ‘apply’ /thîiwâakaan/ ‘office’ /wíphâak/ ‘comment’ /nákkhít/ ‘thinker’ /sǎŋkhró/ ‘synthesise’ /râakthǎan/ ‘foundation’ /níyom/ ‘doctrine’ /ʔùdomkháti/ ‘ideal’ theory /sǎŋkhomsàat/ ‘social science’ /sèetthàsàat/ ‘economics’ /nákcítwítyaa/ ‘psychologist’ /cítwítyaa/ ‘psychology’ /phúunthǎan/ ‘foundation’ social /chəəŋ/ ‘base on’ /ʔaanaaníkhom/ ‘colony’ /sítthímánútsàyáchon/ ‘human right’ /látthí/ ‘doctrine’ /phúthásàatsànǎa/ ‘Buddhism’ /phɔɔphiən/ ‘sufficient’ /yúthhūu/ ‘adhere to’ /ʔàthíbaay/ ‘explain’ /kiàwkàp/ ‘about’ /maanútsàyáwítthàya/ ‘anthropology’ /sǎnǎə/ ‘propose’ /praakòtkaan/ ‘phenomenon’ /tôoyéəŋ/ ‘oppose’ /phaaytây/ ‘under’ /dǎŋdǎəm/ ‘traditional’ /sàthóon/ ‘reflect’ /khroon̄sǎaŋ/ ‘structure’ /sòotkhlóon/ ‘harmonious’ /níyaam/ ‘definition’ /hěndūay/ ‘agree’ /waathá/ ‘word, idea’ (part of a person’s name) /chumchon/ ‘community’ /khàtyéəŋ/ ‘oppose’ /sǎmǎymà/ ‘modern time’ /rírêəm/ ‘originate’ /thópthuan/ ‘review’ /rûppàtham/ ‘concrete’

The collocates of *idea* ranked by Log Ratio score are as follows:

preconcieved cross-fertilisation **half-baked** **new-fangled** **crackpot** nationhood propound **foggy** receptive **soff** moot jot **bereft** innovative expound **fanciful** **mistaken** **hazy** inspirational disseminate common-sense dissemination toy interchange fundraising espouse mathematical **erroneous** diffusion **germ** impart freudian **faint** flirt **reject** **vague** **simplistic** **obsessed** **grandiose** conceive mull plato **opposed** femininity enlightenment **propagate** formulate sharing creativity relish unitary abstract originate convey imaginative innate **absurd** **shrew** stimulating presuppose excite citizenship bliss keynesian entertain embody **scorn** **discredit** darwin **abandon** **daft** maxíst splendid creative marx revive

6.5.1.26 /ʔeekkàchon/ vs. *individual*

The collocates of /ʔeekkàchon/ ranked by Log Ratio score are as follows:

/kɔɔ kɔɔ rɔɔ/ (acronym for ‘the joint standing committee’) /kɔɔ chɔɔ/ (acronym for ‘private education’) /sɔɔ chɔɔ/ (acronym for ‘the private education commission’) /kɔɔ rɔɔ ʔɔɔ/ (acronym for ‘the joint cooperation committee’) /ʔùdomsùksǎa/ ‘higher education’ 2525 private /phâak/ ‘part’ /khánaacaan/ ‘faculty members’ 2546 /ʔonkɔɔn/ ‘organisation’ /ʔonkɔɔnphátthánaaʔeekkàchon/ ‘non-governmental organisation’ /bɔɔríhǎanñaan/ ‘administer’ /sùanthóonhìn/ ‘local’ /sàthǎaban/ ‘institution’ /nùayràatchákaan/ ‘government office’ /hǎanraan/ ‘store’ /kammásit/ ‘ownership’ /ʔaachiiwásùksǎa/ ‘vocational education’ /rátwísǎahàkít/ ‘state-owned enterprise’ /phûuloŋthun/ ‘investor’ /roonpháyaabaan/ ‘hospital’ /rát/ ‘state’ /wâacâaŋ/ ‘employ’ /roonriaan/

‘school’ /phûuthɛn/ ‘representative’ /kaanlonthun/ ‘investment’ /nùayñaan/ ‘institute’ /naaythun/ ‘capitalist’ /sùanrûam/ ‘contribution’ /sòŋsǎəm/ ‘support’ /phûupàtibàtñaan/ ‘officer’ /ʔùppàphòok/ ‘consume’ /sǎmpàthaa/ ‘concession’ /sòŋsǎəmkaanlonthun/ 2547 /ʔonkaan/ ‘organisation’ /ʔaasǎasàmàk/ ‘volunteer’ /hǒkaankháa/ ‘chamber of commerce’ /khlínik/ ‘clinic’ /bǒrísàt/ ‘company’ /thúrákit/ ‘business’ /wísǎahàkit/ ‘enterprise’ /sàthǎan/ ‘place’ /phûuphàlit/ ‘producer’ /lonthun/ ‘invest’ /ʔùtnǔn/ ‘support’ /ʔitsàlaam/ ‘Islam’ /sǎŋkàt/ ‘belong to’ /nákthúrákit/ ‘businessperson’ /rûammuu/ ‘cooperate’ /phraráatchábanyàt/ ‘act’ /phâakwíchaa/ ‘department’ /khwaamrûammuu/ ‘cooperation’ /phûupràkòpkaan/ ‘entrepreneur’ /sǒŋkhró/ ‘support’ /chǎlǒ/ ‘retard’ /prǎsǎannaan/ ‘coordinate’ /khánákammákaan/ ‘committee’ /bǒrípòok/ ‘consume’ /lǎwrían/ ‘study’ /càthǎa/ ‘provide’ /hǎaruu/ ‘consult’ /càttǎn/ ‘found’ /khàyyǎaytua/ ‘expand’ /wítháyaalay/ ‘college’ /kaansùksǎa/ ‘education’ /phòokháa/ ‘merchant’ /bòtbàat/ ‘role’ /sànap̀sànǔn/ ‘support’ /fayfǎa/ ‘electricity’ /phûubǒrìhǎan/ ‘executive’

The collocates of *individual* ranked by Log Ratio score are as follows:

hiv-infected self-fulfilment named like-minded **predispose** **infected** **coerce** **affected** enterprising psychologically identifiable motivated his/her **isolated** uniqueness **susceptible** self-interest motivate wealthy talented interact randomly autonomous reside **deviant** well-being vary empower private self-esteem motivation **selfish** functioning privacy organisation trait antonomy **powerless** subjective unrelated group liberty **constrain** **solitary** collective compose fitness freely competent perception lifestyle personality socially selected counselling conform conception mature donation **marginal** reproduce given lifetime nominate assign unique healthy

6.5.1.27 /naayókrátthàmontrii/ vs. *prime minister*

The collocates of /naayókrátthàmontrii/ ranked by Log Ratio score are as follows:

/lámóət/ (part of a person’s name) /rát/ (part of a person’s name) /chiiwá/ (part of person’s name) /chinnáwát/ (person’s name) 66/2523 /liikphay/ (person’s name) /sò/ (part of a person’s name) /yaará/ (part of a person’s name) /wêet/ (part of a person’s name) /bɛɛ/ (part of a person’s name) /ʔàphisit/ (person’s name) /kǒdðǒn/ (Gordon) /yút/ (part of a person’s name) /banhǎan/ (person’s name) /siŋ/ (part of a person’s name) /sǎmnák/ ‘office’ /phátsàdù/ ‘inventory’ /kray/ (part of a person’s name) /wálaa/ (part of a person’s name) /wǎən/ (part of a person’s name) /rǒ mǒ tǒ/ (acronym for ‘minister’) /hun/ (part of a person’s name) /phíbuunsǒŋkhraam/ (person’s name) /tháksǐn/ (person’s name) /ʔəən/

(part of a person's name) /phoo tsoo thoo/ (acronym for 'Police Lieutenant General') /khoo/ (part of a person's name) /hút/ (part of a person's name) /ʔaachaa/ (part of a person's name) 10300 /ʔaan/ (person's name) /təy/ (part of a person's name) /nəo səo/ (acronym for 'Miss') /rəoŋnaayókráthàmontri/ 'Deputy Prime Minister' /phánáthàn/ (title) /khoosòk/ 'spokesperson' /krápbaŋkhomthuun/ 'inform (the king)' /tin/ (part of a person's name) /cia/ (part of a person's name) /ʔəpdun/ (part of a person's name) /baaw/ (Brown) /cù/ (part of a person's name) /mia/ (part of a person's name) /châatchaay/ (person's name) /kriaj/ (part of a person's name) /krompràchaasāmphān/ 'Public Relations Department' /rāatcháʔoonkaan/ 'king's command' /mɛɛ/ (part of a person's name) /lút/ (part of a person's name) /palət/ 'deputy' /rátthàmontri/ 'minister' /māayhèet/ 'notes' /bɛɛ/ (part of a person's name) /pan/ (part of a person's name) /siw/ (part of a person's name) /lák/ (part of a person's name) /chaa/ (part of a person's name) /sīnlápá/ (part of a person's name) /cháwálit/ (person's name) /non/ (part of a person's name) /ráksāakaan/ 'act for' /rát/ (part of a person's name) /sàwət/ (part of a person's name) /leekhāthikaan/ 'secretary-general' /nəo rəo/ (acronym for 'Office of the Prime Minister') /hūanāaphák/ 'political party leader' /kamphuuchaa/ 'Cambodia' /ʔədiit/ 'former' 2534 /pəo pəo səo/ (acronym for 'Office of Narcotics Control Board') /sen/ (part of a person's name) /khàcəon/ (part of a person's name) /səŋkaan/ 'command' /khánárátthàmontri/ 'cabinet' /khəon/ (part of a person's name) /thàlɛɛŋ/ 'declare' /phonʔèek/ 'General' /kráthūuthāam/ /chun/ (part of a person's name) /rábiap/ 'regulation' /sūnthəon/ (part of a person's name) /thaanin/ (person's name) /pliantua/ 'replace' /kitti/ (part of a person's name) /thii/ (part of a person's name) /khánárátthàbaan/ 'government' /foon/ 'telephone' /rátthàmontriwāakaan/ 'minister' /tháná/ (part of a person's name) /cəomphon/ 'Field Marshal' /mōopmāay/ 'assign' /coow/ (part of a person's name) /thamniap/ 'official residence' /lii/ (part of a person's name) /kàlaahōom/ 'Ministry of Defense' 2534 /phóot/ 'post' /banchaakaan/ 'command' /sēnii/ (part of a person's name) /yút/ (part of a person's name) /yoŋ/ (part of a person's name) /ráppàak/ 'promise' /phíthiipəət/ 'opening ceremony' /mōopʔamnāat/ 'authorise' /cinda/ (part of a person's name) /pàw/ (part of a person's name) /həŋchəop/ 'approve' /preem/ (person's name) /tamnɛɛŋ/ 'position' /damroŋ/ 'take a position of' /prasit/ (person's name) /sūu/ (part of a person's name) /phonʔèek/ 'General' /laaʔòk/ 'resign' /khəo rəo məo/ (acronym for 'cabinet') /phrábəorom/ 'royal' /praasāy/ 'make a speech' /thànəom/ (person's name) /tɛɛŋtāŋ/ 'appoint' /krəphōm/ (pronoun) /wòot/ 'vote' /praamòot/ (person's name) /kráp/ 'prostate oneself' /thábuaj/ 'bureau' /práthaanaathípbəodii/ 'president' /deewit/ (David) /kamkápduulɛɛ/ 'supervise' /loŋnaam/ 'sign' /práthaan/ 'chairman' /khamsəŋ/ 'command' /kəo tsoo rəo/ (acronym for 'Office of the Police Commission') /sàrit/ (person's name) /hoo/ (part of a person's name) /pəo/ (part of a person's name) /pràcam/ 'posted' /thoo/ (part of a person's name) /hāaruu/ 'consult' /paathàkàthāa/ 'speech' /rəon/ (part of a

person's name) /sian/ (part of a person's name) /haa/ (part of a person's name) /laa/ (part of a person's name) /maan/ (part of a person's name) /sǎmnákɲaan/ 'office' /wáywaaŋcaɲ/ 'trust' /phûubanchaakaan/ 'commander in chief' /sǎŋkàt/ 'belong to' /rɔŋ/ 'deputy' /chí/ (part of a person's name) /fǎaykháan/ 'opposition' /yóklâk/ 'cancel' /chuan/ (person's name) /pràsǎannaan/ 'coordinate' /leekhǎanúkaan/ 'secretary' /phûurápphitchɔɔp/ 'person in charge of' /thûiprùksǎa/ 'consultant' /ʔɔɔtreelia/ 'Australia' /pràchaathíppàtay/ 'democracy' /yúp/ 'dissolve' /woŋ/ (part of a person's name) /ʔitsáaraaʔeen/ 'Israel' /pèətpǎəy/ 'reveal' /tɔɔwǎa/ 'blame' /sàphaaphûuthɛnraatsàdɔɔn/ 'House of Representatives' /ʔittaali/ 'Italy' /rátthàsàphaa/ 'parliament' /samàk/ 'apply for' (person's name) /ràatcháakaan/ 'government service' /cuu/ (part of a person's name) 2550 /naykhànánán/ 'at that time' /maaleesia/ 'Malaysia' /khrúkkhrit/ (person's name) /rɔɔ mɔɔ wɔɔ/ (acronym for 'minister') /pràkàat/ 'announce' /yuan/ 'visit' /nûaŋ/ 'owing to' /sàk/ (part of a person's name) /klàaw/ 'say' /phûuchûay/ 'deputy'

The collocates of *prime minister* ranked by Log Ratio score are as follows:

mouloud kazimiera al-zubaydi sidki holkeri zubi kanisese laraki atef leekpai gasan vitol mudar gediminas ghafar dashiyn shapour aneerood lubbers tzannis panyarachun kiichi stoph sibomana mian narasimha ingvar sid-ahmed toshiki anibal cheiffou hamrouche kokou sasagov greguric calfa merbah yildirim vishwanath honorat cavaco itzhak sa'adoun konstantinos wilfried choonhaven fokin hamzah rabbie chok nawaz razanamasy mesut yasuihiro byambasuren mohamad akbulut yitzhak tzannetakis andreotti pratap mitsotakis mahathir poul badran bielecki jugnauth Németh jozsef tadeusz ruud giulio mulroney Schlüter shik noboru khaleda vagnorius oye bildt giuliano markovic antall Miklós azzedine yilmaz benazir gaidar milongo sandiford aho hammadi prasad ghazali karami koffigoh rocard diro goh sigua ryzhkov prunskiene demirel bhattacharya koirala gro amato nakasone casimir brundtland krzysztof tengiz bakhtiar suchocka rao petre namaliu sharif kaifu miyazawa yegor suleyman anand mazowiecki shamir beregovoy abalkin adrien bufi cresson adamec bakr carlsson rajiv umar ratu lukanov seri olszewski taoiseach tshisekedi guerra kiet modrow nguza zaid González rabin shekhar chandra valentin kuan amadou hawke chuan haughey chatichai bolger datuk nikolai pavlov Márquez filip keating alphonse tariq bhutto papandreou meciar dimitrov haider harlem deputy zia ghulam takeshita mohamed suchinda mohammad ante lange Maizière reappoint peng likud begum muang manley chirac saad felipe fabius silva andreas leonid michel rashid mahmud marten mustafa callaghan shaikh aziz lal vere hanna asquith salim tong malaysian belize erskine vaclav vladimir muldoon hun sen etienne attlee klaus muhammad andrei iliescu maj.-gen. outgoing singh slovak viktor thatcher Özal marian pierre walesa gandhi edith ppp gen. kinnock disraeli clement constantine ahmed beleaguered hon.

czechoslovak It.-gen. gladstone -gen caretaker reshuffle macmillan alfonso sergei yew baldwin harold abdul abu elias israeli mba howe resignation eden margaret pitt winston balfour churchill malaysia **assassinate** re-elect cabinet mohammed cpngo pakistani André appoint col. former jan heath reynolds mahmoud li resign hungarian havel vice-president willi maastricht lothar interim nu thai jacques edouard czech transitional mitterrand hans croatian ukrainian estonia mikhail downing khan thailand latvia federal chancellor ivan lebanese danish chamberlain heseltine palmer elect neville juan geoffrey carlos minister salisbury acting nov. mr. singapore coalition syrian appointment **assassination** 1916 ted brian sept. stanley albert pakistan oct. preside refer milan belgian dec. list major kohl bulgaria president adviser canadian Wilson swedish feb. japanese jan. macdonald mitchell nominate turkish announce leader ali hungary foreign **dismiss** affair hitherto australian defence john november prime successor iraqi ballot aug. february v. israel indian lamont visit pledge mps attach tourism jean lloyd summit spanish gorbachev bob commons predecessor irish kingdom conservative

6.5.1.28 /ʔeekkàsãan/ vs. *document*

The collocates of /ʔeekkàsãan/ ranked by Log Ratio score are as follows:

/plɔɔmpleɛŋ/ ‘counterfeit’ 4-01 /pràkùatraakhaa/ bid’ /bayráp/ ‘receipt’ /ʔãaŋʔiŋ/ ‘refer’ /hɔ̀còtmãayhèet/ ‘archive’ /láp/ ‘confidential’ /lãkthãan/ ‘evidence’ /nêep/ ‘enclose’ /nãŋsũudənthaaŋ/ ‘passport’ /sãmnaw/ ‘copy’ /fêem/ ‘file’ /fêek/ ‘fax’ /pùk/ ‘pack’ /prãkòɔpkaaŋ/ /rábiapwaará/ ‘agenda’ /baykhõnsĩnkháa/ ‘bill of lading’ /khũumũu/ ‘manual’ /còtmãayhèet/ ‘annals’ /ʔilëkthoonik/ ‘electronic’ /ʔaayát/ ‘seize’ /thábian/ ‘register’ /loŋchũu/ ‘sign’ /sãmùt/ ‘notebook’ /siŋphim/ ‘printed matter’ /pháyaan/ ‘witness’ /banchii/ ‘account’ /hiiphòɔ/ ‘package’ /kèprãksãa/ ‘keep’ /thét/ ‘untrue’ /khamkhòɔ/ ‘petition’ /sòŋmòɔp/ ‘hand over’ /kròɔk/ ‘fill in’ /phròómthán/ ‘together with’ /sòŋ/ ‘envelope’ /bayrápròŋ/ ‘certificate’ /yũun/ ‘submit’ /loŋthábian/ ‘register’ /thàay/ ‘photocopy’ /khrópthũan/ ‘completely’ /thòykham/ ‘word’ /prãthápra/ ‘affix’ /sàp/ ‘all’ /plɔɔm/ ‘fake’ /cãttham/ ‘make’ /kamkàp/ ‘accompany’ /cãttriam/ ‘prepare’ /piiŋóppràmaan/ ‘fiscal year’ /cèekcàay/ ‘distribute’ /laaymũuchũu/ ‘signature’ /mãaylêek/ ‘number’ /rápròŋ/ ‘certify’ /sìt/ ‘right’ /chànòot/ ‘title deed’ /cèek/ ‘distribute’ /sãmmãnaa/ ‘seminar’ /phoŋsãawádaan/ ‘historical annals’ /rĩapriaŋ/ ‘write’ /sũunhãay/ ‘lost’ /loŋnaam/ ‘sign’ /sen/ ‘sign’ /sũnlákaakòɔn/ ‘customs’ /tõnchàbàp/ ‘manuscript’ /tòɔpaynii/ ‘from now on’ /wátthù/ ‘object’ /khamróŋ/ ‘petition’ /sòŋ/ ‘send’ /chíicèeŋ/ ‘explain’ /phõyphrêe/ ‘disseminate’ /lêekthii/ ‘number’ /cêeŋkhwaam/ ‘inform (the police)’ /khòkhwaaam/ ‘statement’ /kiawkhòŋ/ ‘related to’ /khãakhãw/ ‘inbound’ /phaachána/ ‘container’

/trùatsòp/ ‘check’ /khǒoráp/ /nǐisǐn/ ‘debt’ /pràcamtua/ ‘individual’ /raaylá?iat/ ‘detail’ /raaychûu/ ‘name list’ /phróom/ ‘together with’ /prákòp/ ‘supplement, do’ /khónkhwáa/ ‘study intensively’ /còtmáay/ ‘letter’ /rûapruam/ ‘compile’ /tamraa/ ‘textbook’ /phâaphthàay/ ‘photograph’ /rábù/ ‘specify’ /khômúun/ ‘data’ /waarásáan/ ‘journal’ /bancù/ ‘contain, assign someone to a post’ /khòthétcin/ ‘fact’ /khát/ ‘copy’ /khón/ ‘search’ /tiiphim/ ‘publish’ /kammásit/ ‘ownership’ /khwaamthùuktòu/ ‘justice’ /phim/ ‘publish’

The collocates of *document* ranked by Log Ratio score are as follows:

leaked sgml **interrogatory** workflow unsigned consultative hypertext forged **incriminating** peruse classified imaging sheaf archival confidential signed drafting append **falsify** wordperfect retrieval deed filing retrieve consultation vatican weighty disclosure facsimile draft disclose contractual circulate emanate **leak** tracking **purport** signature printed written registration scanner authentic **execute** ratification interchange despatch electronic copy relevant inspect briefing supporting verification processing scan handwriting archive discovery structured enclosed font tender spreadsheet title sign document **execution** memorandum transmit appellant preparation signing embody publish arbitration **lengthy** prepare format entitle artifact summons embed bundle inspection surviving published print guidance specified outline legally file originate contain relate summarise produce passport **seize** **lodge** formal concerning forge edit legal des charter text inland possession submit original curriculum discussion dip proposal extract endorse publishing content constitutional official translation internal insert incorporate send summary stamp historical graphics

6.5.1.29 /nɛwthaŋ/ vs. way

The collocates of /nɛwthaŋ/ ranked by Log Ratio score are as follows:

/nóomnam/ ‘adopt’ /sàññǎnǎ/ ‘recommend’ /chíiné/ ‘recommend’ /pròuŋdòu/ ‘harmonise’ /kaanphátthánaa/ ‘development’ /kamkàpduulee/ ‘supervise’ /damnəŋŋaan/ ‘operate’ /buuránaakaan/ ‘integration’ /khòwsàññǎnǎ/ ‘recommendation’ /kêkǎhǎy/ ‘solve’ /pàtibàt/ ‘perform’ /phraráatchádamri/ ‘royal idea’ /chíinam/ ‘guide’ /phòophian/ ‘sufficient’ /khûumuu/ ‘manual’ /hǎaruu/ ‘consult’ /lákkeen/ ‘rule’ /thunníyom/ ‘capitalism’ /pàtirûup/ ‘reform’ /yútthûu/ ‘adhere to’ /pràpprun/ ‘improve’ /sǎŋkhomníyom/ ‘socialism’ /khónkhwáa/ ‘study intensively’ /rêŋrát/ ‘speed up’ /sòotkhlóu/ ‘harmonious’ /sàññǎ/ ‘propose’ /phátthánaa/ ‘develop’ /pràchaa/ ‘people’ /sàwɛŋhǎa/ ‘seek’ /kêe/ ‘solve’ /damnəŋ/ ‘manage’ /sàməanchǎn/ ‘united’ /sòəmsâan/ ‘enhance’ /sǎnti/ ‘peace’

/kròp/ ‘framework’ /măw/ ‘presume’ (part of a person’s name) /càttham/ ‘make’ /phráràatcháthaan/ ‘give’ /càtwaan/ ‘arrange’ /náyoobaay/ ‘policy’ /kamnòt/ ‘specify’ /**rápmuu**/ ‘cope with’ /damnəənkaan/ ‘manage’ /yiawyyaa/ ‘treat’ /pàtibàtṭaan/ ‘work’ /mâattràkaan/ ‘measure’ /phlākdan/ ‘push forward’ /yúthásàat/ ‘strategy’ /lākkaan/ ‘principle’

The collocates of *way* ranked by Log Ratio score are as follows:

ickniel wend pave append conceiving fosse pennine sure-fire milky akers tax-efficient principled inimitable **off-hand** **wheelde** behave **grope** **bulldoze** conceptualise matter-of-fact **chomp** whichever **slovenly** cost-effective painless parting **high-handed** indefinable **underhand** tactful time-honoured **haphazard** relieving munch **bluff** communicating oblique cotswold revolutionise navigate plod analogous roundabout weave phrasing meaningful downhill claw **maze** ingenious carriageway **uncanny** winning systematic categorise **devious**

6.5.1.30 /muəŋthay/ vs. *Thailand*

The collocates of /muəŋthay/ ranked by Log Ratio score are as follows:

/nakkaan/ /rókrâak/ ‘homeland’ /prākanchiiwít/ (part of a proper noun) /nɛɛwnâa/ ‘first-class’ /phamnák/ ‘dwell’ /thammaahăakin/ ‘make a living’ /maarian/ /chûudan/ ‘famous’ /lay/ ‘university’ (part of a place name) /chánnam/ ‘first-class’ /rát/ (part of a place name) /klàp/ ‘return’ /ʔùdomsǝmbuun/ ‘abundant’ /hít/ ‘popular’ /thúkwannii/ ‘at present’ /hăakin/ ‘make a living’ /châatthay/ ‘Thailand’ /sèetthī/ ‘rich man’ /**rábàat**/ ‘spread (of disease)’ /fàràŋ/ ‘westerner’ /thiaw/ ‘travel’ /kràman/ ‘probably’

The collocates of *Thailand* ranked by Log Ratio score are as follows:

phuket pattaya laos malaysia philippines brunei singapore bangkok indonesia burma taiwan thai vietnam cambodia asia pacific korea sri lanka exotic gen. pakistan border india **refugee** hong kong asian china export camp japan gulf southern holiday australia visit import prime james western relation northern military foreign

6.5.1.31 /lākthǎan/ vs. *evidence*

The collocates of /lākthǎan/ ranked by Log Ratio score are as follows:

/**camnon**/ ‘surrender’ /pháyaan/ ‘witness’ /mōoránábàt/ ‘death certificate’ /sùuppháyaan/ ‘investigate’ /booraankhádi/ ‘archeology’ /ʔèkkàsǎan/ ‘document’ /bèpphim/ ‘form’ /namsùup/ ‘attest’ /laaymùuchùu/ ‘signature’ /nítiwítháyaasàat/ ‘forensic science’ /**phísùut**/ ‘prove’ /**thét**/ ‘untrue’ /**nêechát**/ ‘clearly’ /sǎmnaw/ ‘copy’ /rùapruam/ ‘compile’ /**ʔawphít**/ ‘implicate’ /baysèt/ ‘receipt’ /bayráp/ ‘receipt’ /**hàkláaŋ**/ ‘confute’ /thábianbâann/ ‘house registration’ /rôŋrôŋ/ ‘trace’ /**cêeŋkhwaam**/ ‘inform (the police)’ /bòsɛɛ/ ‘clue’ /**nêennǎa**/ ‘strong’ /nêep/ ‘enclose’ /khamkhǒw/ ‘petition’ /**klàawhǎa**/ ‘accuse’ /**sòppàakkham**/ ‘interrogate’ /**yunyan**/ ‘assure’ /phrómthán/ ‘together with’ /silaa/ ‘stone’ /ʔaŋʔiŋ/ ‘refer’ /**mǎaycàp**/ ‘arrest warrant’ /loŋchùu/ ‘sign’ /**khrópthûan**/ ‘completely’ /mát/ ‘tie’ /laaylák/ ‘letter’ /pràkòpkaan/ /sòopsüan/ ‘investigate’ /sùupsüan/ ‘investigate’ /**fanđáy**/ ‘reasonable’ /**ʔaayát**/ ‘seize’ /**kèpráksǎa**/ ‘keep’ /phonşǎawádaan/ ‘historical annals’ /chǎwsùu/ ‘hire-purchase’ /sàmùt/ ‘notebook’ /**khòwthétciŋ**/ ‘fact’ /**chátcêeŋ**/ ‘clearly’ /caarúk/ ‘inscribe’ /khamróŋ/ ‘petition’ /pràwàtsàat/ ‘history’ /bantúk/ ‘record’ /**tôoyéŋ**/ ‘oppose’ /trùatsòp/ ‘check’ /lòŋlǎa/ ‘remain’ /nítì/ ‘law’ /**khónphóp**/ ‘discover’ /yùun/ ‘submit’ /**nâachùu**/ ‘believable’ /sòŋmòp/ ‘hand over’ /sǎmnuan/ ‘case’ /**phòŋphǎn**/ ‘indulgent’ /mòpʔamnàat/ ‘authorise’ /praa/ ‘appear’ /banchii/ ‘account’ /thábian/ ‘register’ /pràcamtua/ ‘individual’ /khòŋklaan/ ‘exhibit’ /chíiceeŋ/ ‘explain’ /kùuyuum/ ‘take out a loan’ /tòopayni/ ‘from now on’ /phánákŋaansòpsüan/ ‘inquiry official’ /bèək/ ‘claim expenses’ /trùatphóp/ ‘find’ /praakòt/ ‘appear’ /náksùup/ ‘detective’ /diʔenʔee/ ‘DNA’ /**phòmtəəm**/ ‘increase’ /**chùathùu**/ ‘trust’ /loŋthábian/ ‘register’ /wítháyaasàat/ ‘science’ /phróm/ ‘together with’ /**damnəənkhádi**/ ‘prosecute’ /sàdɛeŋ/ ‘show’ /kèp/ ‘keep’ /yuum/ ‘borrow’ /**phianphòw**/ ‘adequate’ /nǎŋsùu/ ‘document’ /yan/ ‘assure’ /tônchàbàp/ ‘manuscript’ /**phùutòŋhǎa**/ ‘the alleged’ /cǎwnâathî/ ‘officer’ /phâaphthây/ ‘photograph’ /**khonráay**/ ‘criminal’ /sàmàk/ ‘apply for’ /pràcàk/ ‘empirical’ (person’s name) /rápsâap/ ‘acknowledge’ /**sànàpsànũn**/ ‘support’ /tòŋhǎa/

The collocates of *evidence* ranked by Log Ratio score are as follows:

uncorroborated corroborative **circumstantial** adduce anecdotal admissibility admissible **hearsay** **inadmissible** conclusive incontrovertible irrefutable **incriminating** confirmatory epidemiological serological forensic documentary substantiate **damning** documented ample empirical **unsupported** archeological supporting corroborate facie **fragmentary** sworn taking affidavit **inconclusive** unequivocal **shred** prima observational abundant impressionistic **scant** indisputable

insufficient convincing immunological **scanty** cross-examination
 compelling **flimsy** dating efficacy **fabricate** accumulate **overwhelming**
conflicting **confession** refute unearth radiological copious oral
 experimental **depopulation** **implicate** first-hand **disbelieve** ethnographic
 econometric accumulated **prosecution** indirect **contradict** **disprove**
tangible morphological uncover **contrary** **fabrication** persuasive cite
sufficient **contradictory** suggest witness inquest sift pleading factual oath
 glean attest **tenuous** biochemical submission suggestive histological
surviving **inconsistent** **discredit** anthropological transcript scientific fossil
 appellant **criminal** contention hypothesis plentiful clinical historical
inflammation **unreliable** statistical written reliable arousal renal
 published photographic jury submit medical existence weigh trial
supportive assertion gather concerning **absence** marshal **ambiguous**
 accumulation competence **accused** deduce support provide expert
definitive consistent coroner geological **infection** **convict** growing
disturbing furnish identification indicate obtain increasing whatsoever
satisfactory

6.5.1.32 /sápsĩn/ vs. *property*

The collocates of /sápsĩn/ ranked by Log Ratio score are as follows:

/ʔaayát/ ‘seize’ /ŋəənthəw/ ‘wealth’ /khəaythəttàlàat/ ‘sale by auction’
 /ʔoonkammasit/ ‘alienate’ /nĩisĩn/ ‘debt’ /kammasit/ ‘ownership’ /rip/
 ‘confiscate’ /səŋmədɔp/ ‘hand over’ /kèpráksáa/ ‘keep’ /thəukhrəw/ ‘own’
 /panyaa/ ‘wisdom’ /kìawnhəw/ ‘relevant’ /coorákam/ ‘steal’ /camnəw/
 ‘mortgage’ /pháriyaa/ ‘wife’ /phúuyaw/ ‘minor’ /phaará/ ‘burden’ /ʔoon/
 ‘transfer’ /cháycàay/ ‘spend’ /lúknii/ ‘debtor’ /bamrunráksáa/ ‘maintain’
 /khrəwphkhrəw/ ‘own’ /camnam/ ‘pawn’ /ʔùthít/ ‘dedicate’ /pləttphay/
 ‘safe’ /khwaammānkhān/ ‘wealth’ /yút/ ‘confiscate’ /bandaa/ ‘all’ /nii/
 ‘debt’ /khūsəmrót/ ‘spouse’ /banchii/ ‘account’ /muunkhāa/ ‘price’
 /chāwsúu/ ‘hire-purchase’ /thun/ ‘capital’ /mɔɔrádòk/ ‘heritage’ /kəwŋthun/
 ‘fund’ /khwaamsīahāy/ ‘damage’ /yùuciw/ /súksəw/ ‘hide’ /lāap/
 ‘fortune’ /mānkhān/ ‘rich’ /pəkpit/ ‘cover up’ /camnāay/ ‘sell’ /təkpən/
 ‘become’ /ŋəppràmaan/ ‘budget’ /sūunhāy/ ‘lost’ /ʔəsāŋhāarimmásáp/
 ‘immovable property’ /sittí/ ‘right’ /lómłálaay/ ‘bankrupt’ /phúukhāy/
 ‘seller’ /trúatsəw/ ‘check’ /láməw/ ‘violate’ /khəmooy/ ‘steal’ /sīahāy/
 ‘damaged’ /kitcākaan/ ‘business’ /phūsúu/ ‘buyer’ /mān/ ‘engaged’
 /pləw/ ‘rob’ /chótcháy/ ‘compensate’ /chāysəw/ ‘use’ /yúthhūu/ ‘hold’
 /ləw/ ‘greedy’ /cāwnii/ ‘creditor’ /thét/ ‘untrue’ /təktəw/ ‘inherited’
 /yaasəptit/ ‘drug’ /phátsədù/ ‘inventory’ /bāanruan/ ‘house’ /raaykaan/
 ‘list’ /phanlāan/ ‘billion’ /phəəmpuun/ ‘improve’ /sāan/ ‘court’
 /khunnásəmbət/ ‘qualification’ /khammān/ ‘promise’ /bəwricāak/ ‘donate’
 /cāwkhəw/ ‘owner’ /ŋəənsət/ ‘cash’ /rāmruay/ ‘rich’ /sāmii/ ‘husband’

/pràməən/ ‘estimate’ /chamrá/ ‘pay’ /châw/ ‘rent’ /càtkaan/ ‘manage’
 /phûutaay/ ‘dead person’ /raaydây/ ‘income’ /phánákñaancâwnâathîi/
 ‘officer’ /phûutôñghãa/ ‘the alleged’ /pràkanphay/ ‘insurance’
 /khúmkehrəw/ ‘protect’ /khonráay/ ‘criminal’ /tittua/ ‘have something
 with one’ /khámpràkan/ ‘guarantee’ /súukhãay/ ‘trade’ /sëeriiphâap/
 ‘freedom’ /sômseem/ ‘fix’ /sùantua/ ‘personal’ /pàtcèekbùkkhon/
 ‘individual’ /dâysia/ ‘gain and lose’ /sũunsia/ ‘lose’ /sùtcàrit/ ‘honest’

The collocates of *property* ranked by Log Ratio score are as follows:

mortgaged health-giving repossessed leasehold **write-down** **expropriate**
 nationalize **confiscated** immovable freehold 1925 movable restitution
 revaluation **confiscation** settled **confiscate** landed adjoining non-domestic
 owner-occupied **speculator** revalue intangible developer rented insulating
 rateable p&o intellectual **deception** medicinal **stolen** curative antiseptic
expropriation miscellaneous intrinsic repossess **chattel** valuation
 unoccupied adjoin letting ownership relational communal tycoon £60,000
elastic occupier appropriation bequeath **infringe** valuer entity magnate
 repossession **dispose** mechanical vacant **bankrupt** sb halifax conveyance
observable inherit therapeutic 1882 £250,000 matrimonial disposition
 semantic residential **slump** vest tenancy rental terraced optical owner
 conveyancing private refurbish **debtor** olympia **loot** **alienation** brewing
 neighbouring **damage** lease copyright nationalized polymer detached
healing watts **theft** sub-committee situate possession magical burton
belong **trespass** insure trustee rent portfolio boom commercial trust
 purchase ecclesiastical **mortgage** **endanger** licensed **violate** millionaire
 dealing passing seller **explosive** possess beneficiary inheritance value
seizure transfer tenant acquire ltd. **abolition** holding lending amenity
disposal enjoyment sell **steal** auction landlord asset qualify inspect
investment **empty** **exploitation** acoustic taking inherent electrical sale
liberty deed symmetry beneficial thermal preservation **damaging**
exclusive assign molecular surveyor physical syntactic borrower capitalist
destruction attribute assignment purchaser whatsoever tax repair soar
 taxation **lost**

6.5.1.33 /khamšan/ vs. *order*

The collocates of /khamšan/ ranked by Log Ratio score are as follows:

66/2523 /sãnpòkkhrəw/ ‘The Administrative Court’ /ʔùthəw/ ‘appeal’
 /cháybay/ /sãanchántôn/ ‘The Court of First Instance’ /ʔaayát/ ‘seize’
 /phêəkhăw/ ‘withdraw’ /thúlaw/ ‘improve’ /chûafan/ ‘obey’ /sãan/
 ‘court’ /sãmnaw/ ‘copy’ /khôbạnkháp/ ‘regulation’
 /phánákñaanʔaayákaan/ ‘prosecutor’ /khonráykhwaamsãamâat/

‘incompetent person’ /rápsâap/ ‘acknowledge’ /fâafũun/ ‘violate’ /náptèè/ ‘since’ /câwphánákñaan/ ‘official’ /phánákñaancâwnâathii/ ‘officer’ /sãanʔaayaa/ ‘The Criminal Court’ /khamphíphâaksãa/ ‘judgement’ /sãanʔùthoôn/ ‘The Appeal Court’ /thábian/ ‘register’ /loŋthôot/ ‘punish’ /wâanâay/ ‘obedient’ /banchaakaan/ ‘command’ /fóŋ/ ‘sue’ /yóklôak/ ‘cancel’ /ʔáthípbòodii/ ‘director-general’ /banchaa/ ‘command’ /wínitchǎy/ ‘judge’ /dètkhàat/ ‘decisive’ /khàt/ ‘go against’ /hâyʔòok/ /ráŋáp/ ‘stop’ /phûubanʔhápbanchaa/ ‘chief’ /loŋnaam/ ‘sign’ /còthábian/ ‘register’ /ʔayyákaan/ ‘prosecutor’ /tèŋtân/ ‘appoint’ /rábiap/ ‘rule’ /tàktuan/ ‘warn’ /cêeŋ/ ‘notify’ /tâysũan/ ‘inquire’ /khamróŋ/ ‘petition’ /kàkkhǎŋ/ ‘detain’ /prookræm/ ‘programme’ /naayókrátthàmontrii/ ‘prime minister’ /ʔànúyâat/ ‘allow’ /chûakhraaw/ ‘temporary’ /diikaa/ ‘petition’ /klànkroŋ/ ‘think over’ /pràkàat/ ‘announce’ /laaymæuchûu/ ‘signature’ /pàtibàt/ ‘perform’ /kòtkràsuan/ ‘ministerial order’ /ráprâatchákaan/ ‘work in the government service’ /loŋthábian/ ‘register’ /laaylák/ ‘letter’ /wikoncàrit/ ‘insane’ /hâam/ ‘forbid’ /máti/ ‘resolution’ /khánárátthàmontrii/ ‘cabinet’ /khádii/ ‘lawsuit’ /khòokamnòt/ ‘specification’ /plòt/ ‘dismiss’ /thoôn/ ‘withdraw’ /bandaa/ ‘all’ /pòkkhròŋ/ ‘govern’ /yápyán/ ‘stop’ /phíthák/ ‘protect’ /wéntèè/ ‘unless’ /khádiiʔaayaa/ ‘criminal case’ /rátthàmontrii/ ‘minister’ /yúp/ ‘dissolve’ /khrêŋkhrát/ ‘strict’ /khátkháan/ ‘oppose’ /sòŋ/ ‘teach’ /khamkhǎo/ ‘petition’ /phíphâaksãa/ ‘judge’ /sòŋpsũan/ ‘investigate’ /wínay/ ‘discipline’ /yéŋ/ ‘oppose’ /praysàni/ ‘post’ /sàlà/ ‘relinquish’ /taam/ ‘in accordance with’ /khánákammákaan/ ‘committee’ /ràatchákitaanúbèeksãa/ ‘the government gazette’

The collocates of *order* ranked by Log Ratio score are as follows:

cheque/postal hernus **pecking** kriel substituted vlok **committal** alphabetical descending mandamus marching adriaan **disobey** slorc certiorari mendicant winding-up s.i. countermand chronological ascending cistercian **revocation** **peremptory** restraining mail pursuant magnitude postal supervision **bankruptcy** **semblance** teutonic guardianship **deportation** obey benedictine **quash** templar **rescind** chivalry interim knights garter forestall **eviction** interlocutory debit **annul** **probation** further p&p residence prerogative **decreasing** airbus monastic created repossession **wasted** making restoring numerical restore precedence si franciscan batting **backlog** **banning** polynomial **confiscation** charging temporal **revoke** code preservation comply merit standing registrar maximise **defy** speaker ascertain **contravene** issuing **compulsory** minimize 1936 receiving **postponement** conserve cleanliness parte jesuit petitioner rsc facilitate despatch established mr. sequential cheque **prohibition** **chaos** decency uphold amend disclosure compliance bolster re-establish preserve deacon commencement enforce enforcement reinstate possession **defiance** counteract winding **detention** simplify regularity maintain avoid obtain coherence spontaneous **prohibit** **unjust**

policing restoration sealed placing starter court birth thereof jurisdiction
secure law reverse fulfil discharge pending payable

6.5.1.34 /ηóppràmaan/ vs. budget

The collocates of /ηóppràmaan/ ranked by Log Ratio score are as follows:

0702//wɔɔ/ /raaycàay/ ‘expenses’ 2502 /kàatdun/ ‘run deficit’
 /piihóppràmaan/ ‘fiscal year’ /dun/ ‘balance’ /càtsǎn/ ‘allocate’
 /khárúphan/ ‘durable articles’ /cháycaay/ ‘spend’ /wonηəən/ ‘sum’ 2555
 /ʔàtchiit/ ‘encourage someone by giving them money’ /sǎmnák/ ‘office’
 /nɔɔ rɔɔ/ (acronym for ‘Office of the Prime Minister’) /bèək/ ‘claim
 expenses’ /phěenkaan/ ‘plan’ /phɔɔ rɔɔ bɔɔ/ (acronym for ‘act’)
 /buuránaakaan/ ‘integration’ /ʔànúmat/ ‘approve’ 2548 /ηóp/ ‘budget’
 2553 /ciat/ ‘allot’ /phaará/ ‘burden’ /sínpluaŋ/ ‘wasteful’ /phěenŋaan/
 ‘project’ /pàtithin/ ‘calendar’ 2552 /phùukphan/ ‘obligated’ /thótrɔŋ/
 ‘advance money’ 53 /sìŋkòsâaŋ/ ‘building’ /pràmaakaan/ ‘estimate’
 /pràcam/ ‘of’ 2550 /pàtibàtŋaan/ ‘work’ /ʔùtnǔn/ ‘support’ /chótchəy/
 ‘compensate’ /ʔoon/ ‘transfer’ /rêŋrát/ ‘speed up’ 1.6 /càttham/ ‘make’ 55
 /hěenchôp/ ‘approve’ /prəyátti/ ‘amend a motion’ /phráràatchábanyàt/
 ‘act’ /raayráp/ ‘income’ /phěen/ ‘plan’ 2547 /kaankhlaŋ/ ‘finance’
 /phûuʔamnuaykaan/ ‘director’ /lāan/ ‘million’ /phêəmtəəm/ ‘increase’
 /thûmthee/ ‘devote’ /yúthásàat/ ‘strategy’ /nii/ ‘debt’ /càay/ ‘pay’
 /phěendin/ ‘government’ 30,000 /phɔɔ sɔɔ/ (acronym for ‘Buddhist era’)
 /wisāaman/ ‘uncommon’ /sápsin/ ‘wealth’ /sànàpsànǔn/ ‘support’ 2546
 /pràyàt/ ‘thrifty’ /bɔɔríhāan/ ‘administer’ /rátthàbaanklaaŋ/ ‘central
 government’ /càttriām/ ‘prepare’ /ηəən/ ‘money’ /bɔɔríhāanŋaan/
 ‘administer’ /damnəənŋaan/ ‘operate’ /kritsàdiikaa/ ‘decree’
 /mđɔpʔamnâat/ ‘authorise’ /bùkkhlaakɔɔn/ ‘personnel’ /sùanthóŋthìn/
 ‘local’ /kòsâaŋ/ ‘build’ 2549 /phanláan/ ‘billion’ /bàat/ ‘Baht’ /raaykaan/
 ‘list’ 400 /khánákammaathikaan/ ‘commission’ /rāaŋ/ ‘draft’
 /khôcāmkàt/ ‘limitation’ /mùat/ ‘group’ /máhāasāan/ ‘huge’ /sômthóp/
 ‘add’ /camnêek/ ‘classify’ /wínay/ ‘discipline’ /phátsàdù/ ‘inventory’ 2550
 /rátwísāahàkìt/ ‘state-owned enterprise’ /sùanklaaŋ/ ‘centre’ /khroonkaan/
 ‘project’ /tittaam/ ‘follow’ /traymāat/ ‘three months’ /kitcàkaan/ ‘business’
 /khɔɔ rɔɔ mɔɔ/ (acronym for ‘cabinet’) /thûm/ ‘pay large amount’
 /khđɔráp/ /sđòmseem/ ‘fix’ 1.1 /sđmdun/ ‘balance’ /lonŋthun/ ‘invest’ /ηúat/
 ‘installment’ /rāatchákaan/ ‘government service’ 2551 /khamkhđɔ/
 ‘petition’ /phûuthēen/ ‘representative’ /ηəənkûu/ ‘loan’ /mátì/ ‘resolution’
 /rátthàbaan/ ‘government’ 52 /khâacháycaay/ ‘expenses’ /bûaŋtôn/
 ‘initial’ /klaaŋ/ ‘central’ /chótcháy/ ‘compensate’

The collocates of *budget* ranked by Log Ratio score are as follows:

darman omb **cash-limited deficit** 1990-91 1990/91 **shoestring** delegated
 1991/91 1993-94 expansionary 1989/90 balance-of-payment 1992-93
 1992/93 redistributive devolved 1991-92 £24 % m. balanced devolve
 lamont supplementary **cutback** **austerity** congressional fiscal allocated
 r&d surplus revised budgetary departmental projected recurrent draft
 chancellor gp pentagon unveil allocate federal 1990 approve **shortfall**
tight annual financing **slash constraint** 1993 6% allocation 1991 variance
expenditure billion earmark 1992 £6 ec 1909 £30 million unified
promotional forecast bureau eec defence **cut** franc finance approval
 spending norman staffing manpower estimated £1 revenue imf approved
 1994 advertising proposed 1989 legislature senate day-to-day dec. weekly
 1988 balance overall total announce exceed projection republican proposal
 speech neutral vat functional two-thirds estimate tax boost adoption
 taxation **inflation** nhs combined **limited** household nov. adopt bush
reduction shadow contribution saving modest increase **abolish** jan.
 coming present **reduce** percentage economy

6.5.1.35 /náamman/ vs. *oil*

The collocates of /náamman/ ranked by Log Ratio score are as follows:

/paam/ ‘palm’ /nəysäy/ ‘butter’ /mákòok/ ‘olive’ /roonklàn/ ‘refinery’
 /chùaphləəŋ/ ‘fuel’ /phûukháa/ ‘seller’ /hoo/ ‘gasohol’ /ráhəəy/ ‘evaporate’
 /khùtcò/ ‘drill’ /pám/ ‘petrol station’ /thaantàwan/ ‘sunflower’ /khăaypliik/
 ‘retail’ /thùalǎaŋ/ ‘soybean’ /náammankáat/ ‘kerosene’ /kéet/ ‘gas’ /hũun/
 ‘rancid’ oil /yuukhaalíptát/ ‘eucalyptus’ /phaasíisəpphásáamít/ ‘excise tax’
 /thinnəə/ ‘thinner’ /ŋaa/ ‘sesame’ /náammanbensin/ ‘gasoline’ /klàn/
 ‘distill’ /máphráw/ ‘coconut’ /ʔəəy/ ‘oil’ /ram/ ‘rice bran’ /cháloom/
 ‘anoint’ /sədèt/ ‘drain’ /ciaw/ ‘render down’ /kràthá/ ‘pan’ /wíkritkaan/
 ‘crisis’ /náammanlòóləun/ ‘lubricating oil’ /náammandiiseen/ ‘diesel fuel’
 /sàkàt/ ‘distill’ /tàkiaŋ/ ‘lamp’ /ʔimtua/ ‘saturated’ /phòkkháphan/
 ‘commodity’ /pràkòəpʔaahăan/ ‘cook’ /bensin/ ‘gasoline’ /baareen/
 ‘barrel’ /pitroliam/ ‘petroleum’ /khrâap/ ‘stain’ 95 /phêut/ ‘plant’ /lòóləun/
 ‘lubricate’ /lít/ ‘litre’ /khîaw/ ‘simmer’ /yaaŋphaaraa/ ‘rubber plant’ /pəə
 təə thəə/ (acronym for ‘PTT Public Company Limited’) /təəm/ ‘fill up’
 /həəm/ ‘scented, onion’ /thàanhîn/ ‘coal’ /phrikthay/ ‘pepper’ /thəə/
 ‘turpentine, thermal oil’ /diiseen/ ‘diesel’ /khlaŋ/ ‘tank’ /nəəy/ ‘butter’
 /khăyman/ ‘fat’ /lísəŋ/ ‘peanut’ /phàt/ ‘fry’ /chótchəəy/ ‘compensate’
 /kəəŋthun/ ‘fund’ /khâawphôot/ ‘corn’ /nûat/ ‘massage’ /râat/ ‘pour’
 /krəəŋ/ ‘sieve’ /rótbanthúk/ ‘truck’ /tàwanʔəəkklaaŋ/ ‘Middle East’
 /pluaŋ/ ‘waste’ /thòt/ ‘deep fry’ /kràthiam/ ‘garlic’ /yəəm/ ‘oily’
 /sămrrəəŋ/ ‘store up’ /cùm/ ‘dip’ /bəərisùt/ ‘pure’ /raakaa/ ‘price’ /thăŋ/

‘tank’ /**sînpluaŋ**/ ‘wasteful’ /khrêanyon/ ‘engine’ /súanphàsòm/
‘component’ /**rûa**/ ‘leak out’ /málét/ ‘seed’ /soo/ ‘gasohol’ /roonfayfáa/
‘power plant’ /khõnsòn/ ‘transport’ /**phǎnphũan**/ ‘fluctuate’ /mũu/ ‘pig’
/ʔaynáam/ ‘steam’ /**pràyàt**/ ‘thrifty’ /khlúk/ ‘blend’ /náammandip/ ‘crude
oil’ /bay/ ‘biodiesel’ /phàsòm/ ‘mix’ /banthúk/ ‘load’ /thôv/ ‘pipe’ /kàthí/
‘coconut milk’ /nǎaw/ ‘sticky’ /yòt/ ‘drop’ /khôn/ ‘thick’ /**lòolián**/ ‘nourish’
/yò/ ‘add little by little’ /**khàatkhlɛn**/ ‘lack’ /**thùan**/ ‘illegal’ /khonlǎa/
‘remain’ /mansǎmpǎlǎŋ/ ‘cassava’ /sǎmrèrũup/ ‘ready-made’ /**phɛɛŋ**/
‘expensive’ /khuat/ ‘bottle’ /dùut/ ‘absorb’ /sóv/ ‘sauce’ /pruŋ/ ‘flavour’
/phákchii/ ‘coriander’ /lěew/ ‘liquid’ /thua/ ‘bean’ /dùat/ ‘boil’ /sábũu/
‘soap’ /lálaay/ ‘melt’ /kõŋfay/ ‘bonfire’ /klìn/ ‘smell’ /khlũap/ ‘coat’
/phũuphàlit/ ‘producer’ /chóon/ ‘spoon’ /chúp/ ‘soak’ /mákhũathêet/
‘tomato’ /ʔaaràp/ ‘Arab’ /náamplaa/ ‘fish sauce’ /rêe/ ‘mineral’ /dip/ ‘raw’
/namkhâw/ ‘import’ /liin/ ‘Oleen oil, gasoline, Vaseline, polypropylene’
/camnàay/ ‘sell’ /khâaw/ ‘rice’ /yaasũup/ ‘tobacco’ /**phálanŋaan**/ ‘energy’
/yaanphaaháná/ ‘vehicle’ /sèek/ ‘recite an incantation over’ /fãay/ ‘cotton’
/ʔom/ ‘absorb’ /kram/ ‘gram’ /kheemii/ ‘chemical’ /taw/ ‘stove’ /**pũan**/
‘stained with’ /pàrimaan/ ‘quantity’ /**bamrunràksǎa**/ ‘maintain’ /pêeŋ/
‘flour, powder’ /**sàpphákhuŋ**/ ‘properties’ /thaa/ ‘coat’ /tôm/ ‘boil’
/phũusònʔòk/ ‘exporter’ /**ŋənfəv**/ ‘inflation’ /sáp/ ‘absorb’ /yĩhòv/
‘brand’ /náamyaa/ ‘liquid’ /**thóthhɛn**/ ‘substitute’ /chũm/ ‘soaked’ /klua/
‘salt’ /kròt/ ‘acid’ /phǎw/ ‘burn’ /khwâm/ ‘invert’ /bòv/ ‘pool’ /tàp/ ‘liver’
/sámũmphray/ ‘herb’ /sây/ ‘mantle’ /sàthãanii/ ‘station’ /khêmkhôn/
‘concentrated’ /**tòktàm**/ ‘decline’ /mòv/ ‘pot’ /**pràptua**/ ‘adjust’ /tòv/
‘gland’ /hõv/ ‘shell’ /**khàatthun**/ ‘lose one’s capital’ /ròv/ ‘hot’
/kilookram/ ‘kilogram’

The collocates of *oil* ranked by Log Ratio score are as follows:

safflower petrobar cod-liver grapeseed hamud linseed gholamreza
rapeseed quadrupling us\$30 wheatgerm olive groundnut refinery
sunflower valdez pemex heavy-based **spillage** burmah sullom tbsp
us\$1,600 hazelnut castor sesame mobil exxon crawshaw braer kirkuk
lubricate voe tablespoon primrose turpentine tanker alwyn ultramar opec
soya alaskan **glut** frying **slick** aromatherapy amoco vinegar crude suntan
paraffin sauté **spill** polyunsaturated shale lubricant peanut sump marinate
coconut perfumed citrus pipeline **embargo** water-colour peppermint barrel
gas massage exploration imported abdel aromatic bp tsp grease margarine
canvas rig kuwaiti exporter offshore petrochemical anoint moisturising
exporting pastel alaska garlic onshore vegetable fragrant almond walnut
tar lamp onion bpd refine abdullah acrylic burner proven extraction
arabian drilling shetland lasmo kuwait **dependence** teaspoon
nationalization cod clove lotion sardine collage spoonful gallon petroleum
painting palm scented bahrain resin diesel emulsion tonne hydrocarbon
iranian **chopped** gush importer pepper drill essential bonanza state-owned
lavender **burning** iraqi coal pan import penetrating × saudi yolk hamilton

terminal sprinkle boiling **contaminated** fuel drum lemon mustard refined casserole revenue export sheikh butter gulf shipment **hike** **leak** iraq magnate corn heat blend ½ paint reserve fry heating price oz venezuela brent sea **pollution** saucepan petro producer wax **varnish** virgin herb excluding heater soak **stink** 1973 rosemary finely **smear** consortium **seep** cocoa consumption mineral evaporate **fume** **synthetic** installation maxican litre salt arabia industry midnight discovery norwegian shell **crisis** indigenous sauce seed cargo **shortfall** **volatile** **contaminate** salad modified flour libya juice ton tomato vitamin nigeria iran vapour liver **shock** ore production pasta giant haven sketch scent mix platform north extract gauge bake mexico boom **troubled** pour mining spice rub spoon **exploitation** painter stove **chemical** **sewage** streak brush drip supply shipping tank disperse solvent cooking aberdeen engine quantity reservoir soap electricity storage wheat **burn** arctic peel company ink fish **stain** **fat** cook mixture commodity pump marine pint **rape** medium portrait whale equivalent surplus seal bean nut field skin **dependent** **slump** flavour drain **chop** concession bath remaining smell slice worldwide natural

6.5.1.36 /khâarâatchákaan/ vs. *civil servant*

The collocates of /khâarâatchákaan/ ranked by Log Ratio score are as follows:

/phonláruan/ ‘civilain’ /khâarâatchákaantháhãan/ ‘soldier’ /bamnaan/ ‘pension’ /bamnèt/ ‘pension’ /sãaman/ ‘ordinary’ /tùlaakaan/ ‘judge’ /sãnyaabàt/ ‘royal letter of appointment’ /bɔɔrommáwoŋsãanúwoŋ/ ‘member of the royal family’ /khrɔɔŋchîip/ ‘live’ /phûunóoy/ ‘junior’ /ŋɔɔnduan/ ‘salary’ /kɔɔ tɔɔ rɔɔ/ (acronym for ‘the Police Commission’) /lúukcâan/ ‘employee’ /khâarâatchákaanphonláruan/ ‘civilian employee’ /ʔùdomsùksãa/ ‘higher education’ /bùkkhlaakɔɔn/ ‘personnel’ /tamruat/ ‘policeman’ /kháhàbɔɔdii/ ‘rich person’ /thúrá/ ‘business’ /khâachâw/ ‘rent’ /ʔayyákaan/ ‘prosecutor’ /yôokyáyay/ ‘move’ /ráprâatchákaan/ ‘work in the government service’ /phrâarâatchákritsàdiikaa/ ‘royal decree’ /càttàwaa/ ‘fourth’ /nîisín/ ‘debt’ /tèenɔɔn/ ‘appoint’ /pràcamtua/ ‘individual’ /sãŋkàt/ ‘belong to’ /phûupàtibàtɔɔn/ ‘officer’ /thîaphhâw/ ‘equivalent’ /phásòkníkɔɔn/ ‘people’ /ʔɔɔmsáp/ ‘save up’ /phánákɔɔn/ ‘officer’ /khúnnaan/ ‘nobleman’ /náknãŋsùuphim/ ‘journalist’ /krompàamáy/ ‘Department of Forestry’ /kàsian/ ‘retire’ /kàlaahõom/ ‘Ministry of Defense’ /nákkaanmuan/ ‘politician’ /sãanyúttitham/ ‘Court of Justice’ /yót/ ‘rank’ /phôokháa/ ‘merchant’ /kàsianʔaayú/ ‘retire’ /phûubanɔɔkhápbanchaa/ ‘chief’ /phûunan/ /khruu/ ‘teacher’ /rábiap/ ‘regulation’ /nákhúrákít/ ‘businessperson’ /pàtibàtɔɔn/ ‘work’ /wínay/ ‘discipline’ /khunnáwút/ ‘qualification’ /lúan/ ‘promote’ /tamnèen/ ‘position’ /ʔàtraa/ ‘position’ /bancù/ ‘assign someone to a post, contain’ /câwnaay/ ‘boss’ /rátwisãahàkit/ ‘state-owned enterprise’ /kûakuun/

‘assist’ /nîi/ ‘debt’ /damron/ ‘maintain’ /bankhápbanchaa/ ‘command’ /sĩnbon/ ‘bribe’ /sàanthóonhìn/ ‘local’ /?oon/ ‘transfer’ /kóonhthun/ ‘fund’ 2535 /thaaŋkaan/ /pràcam/ ‘permanent, stationed’ /suanphuumíphâak/ ‘provincial’ /phûuson/ ‘holder’ /bât/ ‘card’ /róonriian/ ‘complain’ /râatchákaan/ ‘government service’ /máhãawittháyaalay/ ‘university’ /bòek/ ‘claim expenses’ /sàthãaban/ ‘institution’ /fùk?òprom/ ‘train’ /cùlaalonkoon/ ‘Chulalongkorn’ /phûuyày/ ‘adult’ /bóorihãan/ ‘administer’ /phrãrãatchábanyàt/ ‘act’ /sùksãathíkaan/ ‘Ministry of Education’ /sithí/ ‘right’ /hũamuaŋ/ ‘province’ /rátthasàphaa/ ‘parliament’ /leekhãanúkaan/ ‘secretary’ /khãntãm/ ‘minimum’ /máhãatthay/ ‘Ministry of the Interior’ 2547 /càriyátham/ ‘morality’ /krãsuay/ ‘ministry’ /cãwnãathii/ ‘officer’ /khãtlũak/ ‘select’ /nãkwichaakaan/ ‘scholar’ /khunnãtham/ ‘virtue’ /sũusãt/ ‘honest’ /phũubóorihãan/ ‘executive’ /tòophthẽen/ ‘reward’ /sòopsũan/ ‘investigate’ /sãmnãkŋaan/ ‘office’ /krom/ ‘department’ /tháhãan/ ‘soldier’ /bandaa/ ‘all’ /sũanklaay/ ‘centre’ /phũuthẽen/ ‘representative’

The collocates of *civil servant* ranked by Log Ratio score are as follows:

senior whitehall retired politician diplomat colonial lawyer minister academic treasury ministry salary clerk indian permanent draft retire career department advise leading former top advice **strike** teacher professional official judge

6.5.1.37 /?onkoon/ vs. *company*

The collocates of /?onkoon/ ranked by Log Ratio score are as follows:

/sàanthóonhìn/ ‘local’ /?oo poo thoo/ (acronym for ‘local administration organisation’) /yuuniséep/ ‘UNICEF’ /?onkoonphátthánaa?èekkàchon/ ‘non-governmental organisation’ NGO /?èekkàchon/ ‘individual’ /phõnkamray/ ‘profit’ /thaaráná/ ‘public’ /pòkkhróon/ ‘govern’ /kamkãpduulẽe/ ‘supervise’ /?aasãasãmak/ ‘volunteer’ /phũuthẽen/ ‘representative’ /cãttãŋ/ ‘found’ /sãŋkèetkaan/ ‘observe’ /chumchon/ ‘community’ /hãanrãan/ ‘shop’ /khruakhàay/ ‘network’ /kũsõn/ ‘merit’ organization /rátwísãahãkít/ ‘state-owned enterprise’ /?itsàrà/ ‘freedom’ /kàsèettãkoon/ ‘agriculturist’ /sithímánútsàyáchon/ ‘human right’ /phẽetháyá/ ‘doctor’ /phãhũphaakhii/ ‘multilateral’ /wichãachĩip/ ‘vocation’ /phãaplák/ ‘image’ /bũkkhlaakoon/ ‘personnel’ /sũumuanchon/ ‘mass media’ /sãhãprãchaachãat/ ‘UN’ /pãtcèekbũkkhon/ ‘individual’ /ruamtua/ ‘assemble’ /nũayŋaan/ ‘institute’ /phũubóorihãan/ ‘executive’ /naanaachãat/ ‘international’ /thóonhìn/ ‘local’ /prãsãanŋaan/ ‘coordinate’ /sãthãaban/ ‘institution’ /sãhãphan/ ‘union’ /muunnĩthí/ ‘charity’ /khãpkhlũan/ ‘drive’ /bóorihãan/ ‘administer’ /sãmaakhom/ ‘association’

/nítibùkkhon/ ‘juristic person’ /khêmkhěη/ ‘strong’ /kòotân/ ‘found’ /bòorihãanŋaan/ ‘administer’ /tambon/ ‘sub-district’ /damnəŋŋaan/ ‘operate’ /bòoricàak/ ‘donate’ /khwaamrûammuu/ ‘cooperation’ /rát/ ‘state’ /sàwěηhãa/ ‘seek’ /tuathēη/ ‘representative’ /máhãachon/ ‘the public’ /càtsǎn/ ‘allocate’ /sǒŋ/ ‘monk’ /thamnâathîi/ ‘perform a duty’ /phâak/ ‘part’ /muunfǒy/ ‘waste’ /sǎa/ ‘public benefit, public’ /sùanklaan/ ‘centre’ /sǎŋkàt/ ‘belong to’ /khwaamphróom/ ‘readiness’ /phùukkhàat/ ‘monopolise’ /cùtmûŋmǎay/ ‘aim’ /phanthámít/ ‘alliance’ /sùusǎan/ ‘communicate’

The collocates of *company* ranked by Log Ratio score are as follows:

joint-stock leathersellers petrobra worshipful unquoted drapers clothworker acquiring market-research mercers d’oyly medium-size blue-chip us-owned privatised quoted levant westborough joint-venture locally-based chicago-based state-owned unlisted foreign-owned family-owned owned american-owned listed multinational **insolvent** unregister factoring privately-held british-owned grupo stationer **liquidation** uk-based non-ec conoco sponsoring british-based second-largest unipart merged government-owned wholly-owned pharmaceutical privately-owned headquarter usm state-controlled subsidiary glasgow-based holding 395 medium-sized stockholding insurance diaghilev daimler-benz winding-up repertory multi-national bes capitalisation haulage relocation non-financial shareholding shipping biotechnology massachusetts-based leasing carbide railroad assembled **disqualification** issuing privatize westinghouse edison directorship marconi incorporated **liquidator** lockheed a&r merging unlimited livery **receivership** marketing subcontract itv nationalise reputable daf california-based generating steamship registrar mining exxon electronics high-technology reimburse **dormant** state-run cash-flow cheapside private-sector **liquidate** debenture mail-order distiller logging omnibus gmbh shareholder goldsmith prudential diversity telecoms petrochemical auditor **defraud** transnational chevron consolidated touring **limited** specialise **moribund** carte ceo rentokil **takeover** relocate flotation london-based start-up employing 3m ferranti one-man inc. courtaulds consortia aerospace wellcome unilever subcontractor **bankrupt** tramway massachusetts publishing daimler high-profile buy-out kodak hi-tech shakespeare privatise participating mba nippon tobacco brewing illustrious governance decca japanese hudson registered trading cose emi associated undisclosed audit invest allot repossess gearing nationalized controlling mersey petroleum telecommunication dow manufacturing plc restructure director **insolvency** consortium malaysian congenial nominee beecham **ailing** secretarial offshore flagship sized swedish oil amalgamate ltd profitable pont 1985 platoon reorganise volkswagen unnamed taiwanese private p&o finnish amstrad hq assurance employee pentium non-executive electricity buying capitalise insider whitbread merge exploration rental xerox boeing india logo retailing asset networking dock reuter share

brewery bus payroll ici in-house equity consultancy theatre venture
 leyland oversees ag engineering merger nissan filing propriety uk reseller
 founder creditor railway restructuring **liability** expatriate finance acquire
profitability dti operating specialize publicly r&d

6.5.1.38 /citcay/ vs. *mind*

The collocates of /citcay/ ranked by Log Ratio score are as follows:

/yútñiaw/ ‘count on’ /klòmklaw/ ‘elevate’ /phòõnphêew/ ‘serene’
 /hòõhiaw/ ‘dispirited’ /ʔòõpʔòõmʔaarii/ ‘generous’ /húkhhòõm/ ‘spirited’
 /bòõpchám/ ‘hurt’ /wáawûn/ ‘restless’ /kràcòõttkràcòõŋ/ ‘frightened’
 /híamkriam/ ‘brutal’ /bèòkbaan/ ‘joyful’ /yàapcháa/ ‘vulgar’ /khùnmuua/
 ‘depressed’ /chêemchûun/ ‘joyful’ /cèemsây/ ‘cheerful’ /khàtklaw/
 ‘refine’ mental /biipkhán/ ‘force’ /phòõnsây/ ‘cheerful’ /diinaam/ ‘good’
 /plòõtplòõn/ ‘free from trouble’ /ràaŋkaay/ ‘body’ /lòõlòõm/ ‘instruct’
 /khápkhêep/ ‘narrow-minded’ /nóomnáw/ ‘persuade’ /yâmyêe/
 ‘miserable’ /yûakyen/ ‘serene’ /bànthòõn/ ‘destroy’ /còõcòõ/ ‘concentrate’
 /ràwròõn/ ‘anxious’ /láʔiatʔòõn/ ‘delicate’ /rûm/ ‘hot’ /ʔòõnwây/
 ‘sensitive’ /lâatnúa/ ‘fresh and blood’ /dètdiaw/ ‘determined’ /sumsâap/
 ‘permeate thoroughly’ /thamráay/ ‘hurt’ /kràchàpkràchêen/ ‘energetic’
 /nêewnêe/ ‘determined’ /wànwây/ ‘swayed’ /pùthùchon/ ‘ordinary people’
 /ʔeenʔiaŋ/ ‘sway’ /hòõthùu/ ‘depressed’ /ràaròõn/ ‘cheerful’ /khêmkhêen/
 ‘strong’ /pànpùan/ ‘frantic’ /canloon/ ‘enhance’ /ròõpkuan/ ‘bother’
 /kràthuan/ ‘affect’ /sâwmòõŋ/ ‘sad’ /khròõppam/ ‘overpower’
 /khêenkrèen/ ‘strong’ /yiawyaa/ ‘treat’ /phòõnkhlaay/ ‘relax’ /sàŋòõp/
 ‘peaceful’ /bòõrisùt/ ‘pure’ /wítòõkkaŋwon/ ‘worried’ /chûarâay/ ‘bad’
 /thaarun/ ‘maltreat’ /ʔòõnyoon/ ‘gentle’ /sàtipanyaa/ ‘wisdom’
 /bùkkhálíkkàphâap/ ‘personality’ /núkkhít/ ‘think’ /thùmthee/ ‘devote’
 /ŋòõkŋaam/ ‘thrive’ /wáawèe/ ‘lonely’ /sàphâap/ ‘condition’ /thîiphûn/
 ‘supporter’ /kilèet/ ‘depravity’ /siat/ ‘penetrate, very high’ /fúunfuu/
 ‘restore’ /khwaamkòõdan/ ‘pressure’ /thòõrámaan/ ‘torture’ /lûaŋrúu/
 ‘foresee’ /thòõthêe/ ‘downhearted’ /ʔùppànísây/ ‘habit’ /ʔûafúaa/
 ‘generous’ /ʔòõnʔee/ ‘weak’ /khwaamkhríat/ ‘stress’ /lúk/ ‘deep’
 /kaayyáphâap/ ‘physical’ /mêettaa/ ‘goodwill’ /sùkkhàphâapcít/ ‘mental
 health’ /sàphaawá/ ‘condition’ /kaay/ ‘body’ /ʔòõráhãn/ ‘Buddhist saint’
 /mânkhon/ ‘stable’ /sòõchûun/ ‘lively’ /hòõtrâay/ ‘cruel’ /sàmàtthàphâap/
 ‘capability’ /sâamákkhii/ ‘harmony’ /sũunson/ ‘sublime’ /khunnátham/
 ‘virtue’ /sàʔàat/ ‘clean’ /phaawánaa/ ‘pray’ /nàknêen/ ‘firm’ /piam/ ‘full
 of’ /ʔaarom/ ‘mood’ /naammátham/ ‘abstractness’ /khâwthûn/ ‘access’
 /sàmaathí/ ‘concentration’ /sòõkàpròõk/ ‘dirty’ /sòõmsâan/ ‘enhance’
 /sùatmon/ ‘pray’ /kràthóp/ ‘affect’ /sũun/ ‘centre’ /bòõkphròõŋ/ ‘defect’
 /ŋòõŋaam/ ‘beautiful’ /khunkhâa/ ‘value’ /sàpsõn/ ‘confused’ /khonraw/
 ‘human’ /sòõmdun/ ‘balance’ /khwaamkhàtyêeŋ/ ‘conflict’

The collocates of *mind* ranked by Log Ratio score are as follows:

one-track boggle inquiring uppermost subconscious indelibly enquiring **unhinge unsound unbidden impressionable** agile bear forefront **fester** recess dualism **flit** whirl imprint **tortured** incisive nimble **turmoil** inventive frame blank wonderfully inquisitive gnaw descartes **numb** **muddled devious wander unconscious implant** peace waking conjure reel refresh **troubled preoccupied** fertile flash resolutely **banish** etch conscious **linger** broaden occupy thought thron **suspicious** foremost intellect **distract** dwell concentrate **lingering** vividly **restless** change trained analytical **lurk** soul spring cross cast **confusion** **stray** emotion **blur** lively

6.5.1.39 /mâattràthãan/ vs. *standard*

The collocates of /mâattràthãan/ ranked by Log Ratio score are as follows:

/mɔɔ ʔɔɔ kɔɔ/ (acronym for ‘Thai Industrial Standards Institute’) GMP ISO /khrɔɔŋchîip/ ‘live’ standard standards /bɪaŋbeen/ ‘deviate’ /bayráprɔɔŋ/ ‘certificate’ /ʔoosoon/ ‘ozone’ /canyaaban/ ‘ethics’ /lɔɔ bɔɔ/ (acronym for ‘cubic’) /sǎakon/ ‘international’ /wíchaachîip/ ‘vocation’ /khunnáphâap/ ‘quality’ /khântàm/ ‘minimum’ /khrɔ̄aŋmǎay/ ‘mark’ /phàlittàphan/ ‘product’ /keen/ ‘criterion’ /monláphít/ ‘pollution’ /thɔɔŋkham/ ‘gold’ /khâachàlià/ ‘mean’ /ráprɔɔŋ/ ‘certify’ /ʔànúkammaakaan/ ‘subcommittee’ /kaanbɔɔrikaan/ ‘service’ /ʔùtsǎahákam/ ‘industry’ /khánaacaan/ ‘faculty members’ /sǔnlákaakɔɔn/ ‘customs’ /khwaamthùuktɔɔŋ/ ‘justice’ /khúru/ ‘teacher’ /chaayfàŋ/ ‘coast’ /thátthiam/ ‘equal’ /kamnòt/ ‘specify’ /khunnáwút/ ‘qualification’ /ʔùdomsùksǎa/ ‘higher education’ /khɔ̄sàŋǎǎné/ ‘recommendation’ /khɔ̄kammnòt/ ‘specification’ /sǎathaarǎnásùk/ ‘public health’ /ʔànaamay/ ‘hygiene’ /càríyátham/ ‘morality’ /sǐnkháa/ ‘goods’ /khɔ̄sɔ̄p/ ‘test’ /náamdùum/ ‘drinking water’ international /banthátthãan/ ‘norm’ /kòtkeen/ ‘regulation’ /thótsɔ̄p/ ‘test’ /thiaphthâw/ ‘equivalent’ /khánákammáakaan/ ‘committee’ /mâat/ ‘measure’ /sǐnlátham/ ‘morals’ /sǐŋwêetlɔ̄ɔm/ ‘environment’ /banchii/ ‘account’ /khûapkhum/ ‘control’ /yaanphaaháná/ ‘vehicle’ /lǎksùut/ ‘programme’ /sòɔtkhlɔ̄ŋ/ ‘harmonious’ /sǎmnákŋaan/ ‘office’ /plòɔtphay/ ‘safe’ /bòŋchii/ ‘indicate’ /càtham/ ‘make’ /níti/ ‘law’ /kammáakaan/ ‘committee’ /pràməən/ ‘estimate’ /pàtibàtŋaan/ ‘work’ /pràpprun/ ‘improve’ /pròŋsǎy/ ‘transparent’ /ŋəəntraa/ ‘money’ /bèepphěen/ ‘convention’ /khunnátham/ ‘virtue’ /baykhònsòŋsǐnkháa/ ‘bill of lading’ /kàsèet/ ‘agriculture’ /tǎm/ ‘low’ /kamkàpduuleɛ/ ‘supervise’ /khunnásòmbàt/ ‘qualification’

The collocates of *standard* ranked by Log Ratio score are as follows:

de-facto frs3 ccitt galland bsi ansi 5750 iso bs5750 ieee bearer **lowering** osi exacting asb cleanliness living accounting facto conformance harmonise craftsmanship deluxe auditing hygiene workmanship bs foreword governance x/open punctuality trading edi conform stringent decency **emission** portability refereeing self-discipline sanitation compatibility ethical reporting **declining** excellence accepted nvq calibration attain propriety raising rigorous cuisine minimum improving required **lax** achievable **deteriorating** agreed safety comply compliance competence maintain high tightening certification attainment affluence uphold frs **deviate** strict nutritional improved audit acceptable measurable adhere insulation housekeeping broadcasting improve accordance recognised applicable les accessibility moral specification internationally adherence lamp enforce **lower deviation** sql finalise exceptionally occupational gold uniformity educational qualifying **effluent** quality literacy decent environmental norm magic technical morality professional raise revised double nationally recommended ec maintenance interface **mandatory** accommodation set proprietary academic hockey **sewage** advertising fairness binary **imposition** reliability accuracy performance satisfactory evening multimedia international ensure reasonable institute telecommunication prescribe monitor nursing audio **tough** meet officer achieve gauge **appalling** consistently accountability uniform charter comparable drinking improvement integrity adoption safeguard specify draft graphics efficiency regulatory **pollution decline** routine approved consistent disclosure monitoring conduct relaxation **impose** absolute proof exceed implement enforcement

6.5.1.40 /nákrian/ vs. *student*

The collocates of /nákrian/ ranked by Log Ratio score are as follows:

/ʔaachiiwá/ ‘vocation’ /mátthàyomsùksää/ ‘secondary school’ /mátthayom/ ‘middle’ /khrûaŋbêep/ ‘uniform’ /naayrua/ ‘navy’ /náksùksää/ ‘university student’ /nísit/ ‘university student’ /pràthömsùksää/ ‘primary school’ /thátsánásùksää/ ‘field trip’ /pràthöm/ ‘primary’ /phûuduuleε/ ‘supervisor’ /mɔɔ/ (acronym for ‘secondary education’) /phûupòkkhrɔɔŋ/ ‘guardian’ /kùatwíchaa/ ‘cram for an examination’ /rooŋrian/ ‘school’ /diidèn/ ‘excellent’ /yúwá/ ‘young’ /hōŋrian/ ‘classroom’ /krèet/ ‘grade’ /ʔànúbaan/ ‘kindergarten’ /kaanriankaansōŋ/ ‘schooling’ /maarian/ /sámrit/ ‘achieve’ /khruu/ ‘teacher’ /pɔɔ/ (acronym for ‘primary education’) /ʔàttraasùan/ ‘proportion’ /lâwrian/ ‘study’ /ʔaachiiwásùksää/ ‘vocational education’ /sãathít/ ‘demonstration’ /piikaansùksää/ ‘academic year’ /**phíkaan**/

‘disabled’ /thun/ ‘scholarship’ /kon/ ‘mechanic, trick’ /wiisâa/ ‘visa’ /hǒp hák/ ‘dormitory’ /phûurian/ ‘learner’ /khǒsǒp/ ‘test’ /loŋthábian/ ‘register’ /chút/ ‘suit (of clothes), group’ /khátlúak/ ‘select’ /chán/ ‘level’ /penrǒŋ/ /rian/ ‘study’ /pràthaan/ ‘chairman’ /bòtrian/ ‘lesson’ /thəəm/ ‘term’ /kàlaahǒom/ ‘Ministry of Defense’ /sàmùt/ ‘notebook’ /ʔùtnǔn/ ‘support’ /fùkhàt/ ‘practise’ /tàaŋchâat/ ‘foreign’ /hay/ ‘high school’ /sǒp/ ‘take a test’ /sǎŋkàt/ ‘belong to’ /raaychûu/ ‘name list’ /cèek/ ‘distribute’ /phrúttikam/ ‘behaviour’ /phûunán/ /tǒntôn/ ‘primary, beginning part’ /sàmàk/ ‘apply for’

The collocates of *student* ranked by Log Ratio score are as follows:

final-year third-year first-year high-school enrolled fourth-year pgce sorbonne post-graduate sit-in second-year enrol semester mature campus phd a-level workbook doctoral sixth-form undergraduate enrolment campus fill-time bursary studentship top-up dissertation beijing jordanstown pro-democracy graduate dormitory humanities tutorial 18-year-old funded **wastage** somerville polytechnic part-time 19-year-old 20-year-old tutor seminary **unrest** **protestor** overseas physics cafeteria his/her graduation placement 22-year-old university msc proficiency tuition lecturer motivated college cohort diploma nu **hardship** faculty fellow advanced 5000 prospective learning eligible vacation qualified fe trainee registered academic scholarship demonstration 7,000 3,000 intake module immerse loan supervisor theological teacher medical fundamentalist supervise vocational introductory entitlement teach cater alike demonstrator chaplain 4,000 encourage recruitment beginner undertake intermediate recruit two-year feedback **fee** counsellor accommodation residence counselling motivate **dissident** entrant ’activist nurse lecture elementary oxford linguistics **protest** guiding instructor **riot** textbook staff enable qualification nursing attend hostel questionnaire honour liaison theology non examination degree assignment chinese

6.5.1.41 /khwaamsăamâat/ vs. *ability*

The collocates of /khwaamsăamâat/ ranked by Log Ratio score are as follows:

capability /khiiit/ ‘limit’ /yǒn/ ‘slacken’ /**khonráykhwaamsăamâat**/ ‘incompetent person’ /**sùamsiá**/ ‘damage one’s reputation’ /chápótua/ ‘individual’ /sàtipanyaa/ ‘wisdom’ /khwaamrúu/ ‘knowledge’ /tháksà/ ‘skill’ /pǒonsàwǎn/ ‘gift’ /**khiiitcamkàt**/ ‘limit’ /sámǔan/ ‘similar’ /thátthiam/ ‘equal’ /khèenkhǎn/ ‘compete’ /**sàkkâyáphâap**/ ‘potential’ /ráy/ ‘without’ /ʔùsăahà/ ‘industrious’ /chàlĭawchàlàat/ ‘intelligent’ /sùttèe/ ‘up to’ /prǎptua/ ‘adjust’ /pràməən/ ‘estimate’ /chĭawchaan/ ‘expert’ /bùkkhálĭkkâphâap/ ‘personality’ /yêkyé/ ‘distinguish’ /tháathaay/

‘challenge’ /sǎəmsâan/ ‘enhance’ /phêəmpuun/ ‘improve’ /thànàt/ ‘apt’ /phûurian/ ‘learner’ /chamnaan/ ‘skilled’ /sàmatthàná/ ‘capacity’ /dǎəy/ ‘inferior’ /phûuyaw/ ‘minor’ /khunnátham/ ‘virtue’ /thân/ ‘amazed’ /diidèn/ ‘excellent’ /khânyùukàp/ ‘depend on’ /sápsǎən/ ‘complicated’ /tǎəpsànǎəy/ ‘fulfil’ /phisèet/ ‘special’ /sùt/ ‘end’ /sûusàt/ ‘honest’ /pràphrút/ ‘behave’ /khánítsàat/ ‘maths’ /wikoncàrit/ ‘insane’ /bòkphrǎəy/ ‘defect’ /pràsòpkaan/ ‘experience’ /sûusăan/ ‘communicate’ /duuthùuk/ ‘look down on’ /bùkkhlaakǎən/ ‘personnel’ /sìnlápin/ ‘artist’ /klân/ ‘hold back’ /thûmthee/ ‘devote’ /phisùut/ ‘prove’ /pràcàk/ ‘evident, realise, empirical’ /khâwthǎn/ ‘access’ /thótsǎəp/ ‘test’

The collocates of *ability* ranked by Log Ratio score are as follows:

aptitude **uncanny** innate **impair** predictive proven linguistic
commensurate willingness communicate withstand cognitive solving
spatial undoubted **discriminate** **overestimate** mobilise verbal athletic
outstrip **manipulate** communicative **impaired** all-round empathy cope
confidence interpersonal flair intellectual mimic talent adapt
underestimate mixed dedication exceptional acting stamina demonstrate
enhance confident comprehend inhibit mathematical inclination utmost
harness artistic organisational regardless innovative possess absorb
declining musical skill attainment **unsure** articulate informed assess
creative enhanced exert evaluate impressed compete heal tolerate
analytical pupil motivation handle **restrict** learner varying amazing mental
reproduce superior correlate remarkable interact listening depend
academic motivate competence technical stimulate independently listener
impress repay retain **undermine** **lesser** respond perceive predict **doubt**
outstanding convey sustain admire conceive **reduced** maturity inherent
logical stimulus develop detect extraordinary adjust personality test
severely **lack** improve acquire

6.5.1.42 /mùubâan/ vs. *village*

The collocates of /mùubâan/ ranked by Log Ratio score are as follows:

/ʔǎə sǎə mǎə/ (acronym for ‘public health volunteer’) village /phâakklaan/
‘central region’ /noon/ (part of a place name) /tambon/ ‘sub-district’
/khǎəyçham/ ‘grocery’ /phâaknǎə/ ‘northern region’ /khôok/ (part of a
place name) /lûukbâan/ ‘villager’ /lǎŋkhaaruan/ ‘household’ /phâakʔiisăan/
‘north eastern region’ /khǎəmmuun/ ‘commune’ /chǎəyŋnǎə/ ‘north east’
/chumchon/ ‘community’ /kàrian/ ‘Karens’ /kiŋʔamphǎə/ ‘sub-district’
/níwêet/ (part of a place name) /chaawkhâw/ ‘hill tribe’ /binthábàat/
‘receive food offerings’ /thaan̄khâw/ ‘entry’ /phâaktây/ ‘southern region’
/kay/ (part of a place name) /hàttàkam/ ‘handicraft’ /chonnábòt/

‘countryside’ /phûuyâyâan/ ‘village headman’ /nǒŋ/ (part of a place name) ‘swamp’ /pràpaa/ ‘water supply’ /khonnay/ /kɔŋthun/ ‘fund’ /sùkkhãaphíbaan/ ‘sanitation’ /ʔaasãasãmàk/ ‘volunteer’ /càtsǎn/ ‘allocate’ /ʔamphəə/ ‘district’ /chəŋkhǎw/ ‘foothills’ /dɔy/ ‘mountain’ /ʔaan̄kèpnáam/ ‘reservoir’ /chaawbâan/ ‘villager’ /kilo/ ‘kilometre’ /hàan̄klay/ ‘far away’ /lávêek/ ‘vicinity’ /ʔiisǎan/ ‘north eastern region’ /bun̄/ ‘marsh’ (part of a place name) /hùpkhǎw/ ‘valley’ /don̄/ ‘jungle’ (part of a place name) /chaaythálee/ ‘coast’ /thêetsàbaan/ ‘municipality’ /phûuthâw/ ‘old man’ /tuayàan̄/ ‘model, example’ /thin̄hǎan/ ‘hometown’ /ráan̄/ ‘deserted’ /róyylá/ ‘percent’ /ʔɔmsáp/ ‘save up’ /ʔùdomsǒmbuun/ ‘abundant’ /bâanruan/ ‘house’ /chaawnaa/ ‘farmer’ /khruaruan/ ‘household’ /sàkkàyáphâap/ ‘potential’ /rápcâan̄/ ‘work for hire’ /thun̄/ ‘field’ /klâykhian̄/ ‘nearby’ /khruakhâay/ ‘network’ /pràmon̄/ ‘fishery’ /kàsèttákam/ ‘agriculture’ /plôn̄/ ‘rob’ /phuukhǎw/ ‘mountain’ /ran̄/ (part of a place name) /dɔn/ ‘highland’ (part of a village name) /thamnaa/ ‘do rice farming’

The collocates of *village* ranked by Log Ratio score are as follows:

tented panmunjom outlying picturesque greenwich farmstead **deserted** unspoiled headman **raze** cotswold stone-built outskirts hamlet hilltop quaint sleepy catterick populous thatched fishing **knockout** lakeside blacksmith surrounding bustling 7:30pm **scattered** neighbouring schoolmaster fete dormitory tranquil hall nestle cotswolds seaside bedfordshire self-sufficient middleton mining inhabitant idyllic tunstall hezbollah nearby wold situate medieval cambridgeshire **idiot abandoned** town charming armenian wooded squire tiny coastal windmill villager orchard moorland commuter oasis northamptonshire colliery shopkeeper leicestershire inn delightful marina **gossip** kurdish burton tribal hillside **remote jumble** wiltshire near postman cornish peaceful downstream somerset churchyard pretty north-west amenity green kilometre pub bobby **rural** feast evacuate south-west **isolated** berkshire footpath dale association riding pit bypass cottage countless communal suffolk saxon bungalow village maiden **winding** moslem overlook mile sutton farming Oxfordshire norfolk farm pleasant global surround planned remains devon dancing olympic shop mountain yorkshire boast cluster antique suburb parish hill historic attractive hampshire manor tribe

6.5.1.43 /sàthǎanthîi/ vs. *place*

The collocates of /sàthǎanthîi/ ranked by Log Ratio score are as follows:

gallery /yòncay/ ‘relax’ /kheehàsàthǎan/ ‘dwelling’ /chániiyá/ ‘sacred place/object’ (the four holy places of Buddhism) /tàakʔaakàat/ ‘go to a

summer resort' /yaanphaahàná/ 'vehicle' /càttriám/ 'prepare' /thátsànáksáa/ 'field trip' /thàaytham/ 'film' /càtlíán/ 'catering' /kàkkhǎŋ/ 'detain' /sàatsànásthàan/ 'religious place' /thòŋthíaw/ 'travel' /thòŋthíi/ 'local' /sàksìt/ 'holy' /càtŋaan/ 'organise (an event)' /nátmǎay/ 'arrange an appointment' /càtsàan/ 'build' /kèpráksáa/ 'keep' /ʔaakhaan/ 'building' /hanniimuun/ 'honeymoon' /wanweelaa/ 'time' /sàñamkiilaa/ 'sports stadium' /booraanwátthù/ 'relics' /ʔànúsòŋ/ 'commemoration' /ná/ 'at' /phíthiipèet/ 'opening ceremony' /phuumlamnaw/ 'hometown' /phákphòŋ/ 'rest' /phátsàdù/ 'inventory' /khápkhêep/ 'narrow' /booraansàthàan/ 'ancient remains' /kaanweelaa/ 'time' /pliaw/ 'out-of-the-way' /kamcàt/ 'get rid of' /thíitǎŋ/ 'location' /ʔeʔàt/ 'congested' /bùkrúk/ 'trespass' /khúnkhəy/ 'familiar' /muunfǔoy/ 'waste' /pàtibáttham/ 'practise the Dharma' /ʔamnuay/ 'facilitate' /nêen/ 'compression' /hàaŋklay/ 'far away' /khón/ 'search' /yòokyáay/ 'move' /sǎathaaràná/ 'public' /tòktèŋ/ 'decorate' /phráràatcháwan/ 'palace' /kòosàan/ 'build' /chaayhàat/ 'beach' /camnàay/ 'sell' : /ràatchákaan/ 'government service' /pràkòpkaan/ /ráankháa/ 'shop' /plòotphay/ 'safe' /nòk/ 'outside' /thíithamkaan/ 'office' /càthǎa/ 'provide' /sìŋkhòŋ/ 'thing' /khlúanyáay/ 'move'

The collocates of *place* ranked by Log Ratio score are as follows:

rillington trysting 20-26 6dj lauriston hiding resting **out-of-the-way** rightful brunswick moorgate assisted portland faraway beauchamp first-team decimal watering n1 play-off langham stopping divers quarter-final far-flung habitable clinch take airy **cremation** articulation worship **hateful** **burial** pilgrimage **usurp** honoured fortified secluded enchanted pride uefa shady nesting **strife** residence **prohibited** concorde congregate **inaccessible** **desolate** semi-final intercourse grosvenor **dreary** idyllic vie **stink** **relegate** 24th exotic mating parking 66 **occupy** sacred holy tranquil regain **lonely** gathering ryder welcoming nursery

6.5.1.44 /chaawbâan/ vs. *villager*

The collocates of /chaawbâan/ ranked by Log Ratio score are as follows:

/khîipàak/ 'gossip' /rácan/ (part of a place name) /phuumpanyaa/ 'intelligence' /chòŋkhǎw/ 'mountain pass' /lávêek/ 'vicinity' /pràat/ 'learned man' /tèektuun/ 'stampede' /nòŋ/ (part of a place name) /hèe/ 'flock in, parade' /kràbuŋ/ 'bamboo basket' /plùkrádom/ 'arouse' /thamnaa/ 'do rice farming' /muŋ/ 'crowd around' /bùkrúk/ 'trespass' /phúuyàybâan/ 'village headman' /ruamtua/ 'assemble' /dùatròŋ/ 'in trouble' /yâakráy/ 'poor' /ninthaa/ 'gossip' /mùubâan/ 'village' /chaaythálee/ 'coast' /sàmátchaa/ 'assembly' /thammaaháakin/ 'make a

living' /**phàwǎa**/ 'frightened' /wíthĩichiiwít/ 'lifestyle' /phraan/ 'hunter' /cèekcàay/ 'distribute' /**léen**/ 'drought' /khàan/ 'dam' /thèep/ 'zone' /ʔòppháyóp/ 'emigrate' /nápthǔu/ 'respect' /**yâakcon**/ 'poor' /panyaachon/ 'elite' /chákchuan/ 'persuade' /**raŋkæ**/ 'bully' /**lòokluan**/ 'deceive' /khǒŋkin/ 'food' /tambon/ 'sub-district' /**thambun**/ 'make merit' /khùt/ 'dig' /**rǒŋrian**/ 'complain' /phâakʔiisǎan/ 'north eastern region' /tèedǎem/ 'at first' /**ʔawpriap**/ 'take advantage of' /phúunbâan/ 'local' /thammádaa/ 'ordinary' /khǒŋ/ 'Mekong' /**rum**/ 'make a combined attack, crowd around' /chumnum/ 'congregate' /wua/ 'cow' /phíthiikam/ 'ritual' /níthaan/ 'tale' /bâanruan/ 'house' /cèek/ 'distribute' /**hǎakin**/ 'make a living' /phôokháa/ 'merchant' /khruaruan/ 'household' /bòo/ 'pool' (part of a place name)

The collocates of *villager* ranked by Log Ratio score are as follow:

flee nearby / **celebrate** crowd village cottage **angry protest** fellow attend **kill** among '

6.5.1.45 /kitcàkam/ vs. *activity*

The collocates of /kitcàkam/ ranked by Log Ratio are as follows:

/nanthánaakaan/ 'recreation' activities /klaançêen/ 'outdoors' /hǒŋrian/ 'classroom' /thátsànásùksǎa/ 'field trip' /phûurian/ 'learner' /kaachâat/ 'red cross' /**kaanriankaansǒn**/ 'schooling' /yúwá/ 'young' /damnǎen/ 'manage' /ʔàdirèek/ 'hobby' /**phrákiat**/ 'honour' /**phanthákǒorǎnii**/ 'obligation' /**chàlǎemchàlǎŋ**/ 'celebrate' /càt/ 'organise' /bùtlǎan/ 'descendent' /kiawnûan/ 'relevant' /**sínlápawátthánátham**/ 'art and culture' /**bambàt**/ 'cure' /kiilaa/ 'sport' /pràchaasǎmphan/ 'public relations' /**ronnáron**/ 'campaign' /lûuksǎa/ 'boy scout' /rûam/ 'join' /**suanrûam**/ 'contribution' /**bamphen**/ 'practise' /khâay/ 'camp' /**sànùksànǎan**/ 'have fun' /làksùut/ 'programme' /làaklǎay/ 'various' /**sòŋsǎem**/ 'support' /**sǎemsǎan**/ 'enhance' /**sǎem**/ 'enhance' /yawwáchon/ 'youth' /**chàlǎem**/ 'celebrate' /wíthháyaakǒon/ 'lecturer' /**banthǎen**/ 'entertainment' /phóppá/ 'meet' /ʔaathí/ 'such as' /sèetthàkit/ 'economy' /**kaanbǒorikaan**/ 'service' /**kùsǒn**/ 'merit' /khròŋkaan/ 'project' /**phlǎetphlǎen**/ 'enjoy' /ruamtua/ 'assemble' /wâaynáam/ 'swim' /sũuŋʔaayú/ 'elderly' /phǎenŋaan/ 'project' /phíthiikam/ 'ritual' /chumchon/ 'community'

The collocates of *activity* ranked by Log Ratio score are as follows:

extracurricular leisure-time spare-time **anti-state** lipase peroxidase out-of-school **pressure-group** **non-core** transcriptional protease **nefarious** sunspot igneous volcanic **frenetic** flurry phosphatase hive promotional

catalytic recreational **subversive** metabolic microbial fund-raising problem-solving kinase purposeful entrepreneurial feverish outdoor **seismic** purposive upsurge modulate **frenzied** leisure sporting **anti-government** neural ceaseless enzyme electrical communicative collaborative cortical pleasurable **curtail** **counter-revolutionary** coordinate **espionage** related governmental cross-border **terrorist** engage intensified curricular **circumscribe** aerobic endogenous **hectic delinquent** modulation raising solving heightened coordinated **displacement** assay inhibit grassroots sphere resurgence **illegal** guerrilla wacc sexual serum vitro strenuous daytime mountaineering **antigen** diplomatic tectonic day-to-day classroom correlate **fraudulent** preventive nocturnal specialized **frantic** organized hub checklist detectable stimulating buzz **covert** undertake regulate cessation spontaneous plasma missionary alkaline ongoing stimulate enjoyable participate everyday platelet incidental **time-consuming** curb productive intense colonic increased **criminal** motor economic innovative bacterial organised co-ordination exploration bustle r&d creative vivo charitable marketing sensory varied pr binding solar sustained structured residual golf associated oversee cellular **prohibit** **disease** muscular **bacterium** corporate **suppression** **interfere** departmental lawful fishing artistic **inhibition** **risky** physical militant supervise cultural **indulge** **speculative** **abnormal** **explosive** functional listening diverse worthwhile meaningful communal **decrease** commercial **incompatible** organizational **reduced** focus **takeover** mining mainstream offshore activate pursue monitor range **restrict** expand enhance biological political participation involved core educational homosexual human integrate competing **confine** energetic encompass organise recreation

6.5.1.46 /bòtbàat/ vs. *role*

The collocates of /bòtbàat/ ranked by Log Ratio score are as follows:

role /phûusàdɛɛŋ/ ‘performer’ /ʔoŋkɔɔnphátthánaaʔèkkàchon/ ‘non-governmental organisation’ /sàthāanápĥāap/ ‘status’ /sàatsànácàk/ ‘religious ruling authority’ /sāmĥan/ ‘important’ /yúti/ ‘end’ /dòotdèn/ ‘distinguished’ /khèɛŋkhǎn/ ‘compete’ /sùumuanĥon/ ‘mass media’ /sūam/ ‘act’ /sàtrii/ ‘female’ /khonklaan/ ‘middleman’ /chínam/ ‘guide’ /dèrchát/ ‘clearly’ /tualákhɔɔn/ ‘character’ /nāathii/ ‘duty’ /panyaachon/ ‘elite’ /phrásǒŋ/ ‘monk’ /tuaʔèek/ ‘leading character’ /pràchan/ ‘compete’ /dèn/ ‘outstanding’ /phûulên/ ‘player’ /khâatwǎŋ/ ‘hope’ /khwaamráppĥitchôp/ ‘responsibility’ /phrámahākàsàt/ ‘great king’ /sǒŋkhraamyen/ ‘cold war’ /weethii/ ‘stage’ /khwaamsāmĥan/ ‘importance’ /kaanmuan/ ‘politics’ /sàthāaná/ ‘status’ /rát/ ‘state’ /khàyǎay/ ‘expand’ /ʔèkkàchon/ ‘individual’ /nákŵícay/ ‘researcher’

/ʔeechiatàwanʔòkchĩaŋtây/ ‘Southeast Asia’ /phêet/ ‘gender’ /rúk/ ‘approach’

The collocates of *role* ranked by Log Ratio score are as follows:

starring mediating pivotal walk-on gatekeeper facilitating enabling reprise **usurp** fulfill underplay photogenesis cameo **interventionist** facilitator **subservient** supervisory play peacekeeping redefinition supporting proactive **reversal** parenting **pathogenic** dual coordinating vis-a-vis mediator housewife legitimise vital crucial leading iran-contrà gender relinquish expanded strengthening redefine **passive** determining breadwinner sweeper watchdog reassessment reappraisal participatory non-executive shaping **downgrade** developmental allot midfield assign policy-making prominent active caring arbiter managerial enhanced **ambivalent** protector cinderella reassess **attacking** rethink dominant feminine audition elucidate decisive changing ascribe respective complementary physiological causal key pioneering caretaker intermediary predominant advocacy constructive advisory consultative **subordinate** reactive clarification important interpreter emphasize challenging therapist counsellor maternal explanatory supportive nato coordinator regulatory functional changed judiciary differing negotiating symbolic clarify parental purchasing perform carer decision-making auditor entrepreneur therapeutic assume influential traditional diplomacy liaison leadership strengthen elected guiding central acting presenter **minor** inspectorate simulation significant **revert** tutor reproductive practitioner learner conception **bureaucracy** define cast major teacher participant protective strategic model syntactic **defensive**

6.5.1.47 /khunnáphâap/ vs. *quality*

The collocates of /khunnáphâap/ ranked by Log Ratio are as follows:

/sǎw məw sǎw/ (acronym for ‘Assessing and Accrediting the Quality of Educational Institutions’) quality /náamtâydin/ ‘underground water’ /sǐnsáp/ ‘asset’ /dǎw/ ‘inferior’ /náamdii/ /líknáy/ ‘lignite’ /sǐŋwêetlǎw/ ‘environment’ /chəəŋ/ ‘base on’ /mâattràthǎan/ ‘standard’ /kaarantii/ ‘guarantee’ /nákwiçay/ ‘researcher’ /phûurian/ ‘learner’ /sùamsoom/ ‘declining’ /chaayfāŋ/ ‘coast’ /thûathǎn/ ‘thoroughly’ /wíçay/ ‘research’ /pàrimaan/ ‘quantity’ /thátthiam/ ‘equal’ /rápprakan/ ‘guarantee’ /phûudǎwʔookàat/ ‘underprivileged’ /thonthaan/ ‘durable’ /sùam/ ‘deteriorate’ /kaanbǔwǐkaan/ ‘service’ /pràpprun/ ‘improve’ /prakan/ ‘guarantee’ /ráprǔwǎŋ/ ‘certify’ /thàanhǐn/ ‘coal’ /sǎnsǎəm/ ‘support’ /kháp/ ‘tight’ /truatšǎw/ ‘check’ /phǎnhàlìt/ ‘product’ /bandit/ ‘graduate’ /chúaphlǎwŋ/ ‘fuel’ /pràməən/ ‘estimate’ /bayráprǔwǎŋ/ ‘certificate’

/wátthùdìp/ ‘raw material’ /lûmnáam/ ‘river basin’ /bamrunràksǎa/ ‘maintain’ /sàwàt/ ‘safety’ /ʔaachiiwásùksǎa/ ‘vocational education’ /krèet/ ‘grade’ /chiiwít/ ‘life’ /khǒŋdii/ /sǐnkháa/ ‘goods’ /keen/ ‘criterion’ /**monláphít**/ ‘pollution’ /phàlittàphan/ ‘product’ /sàpphákhum/ ‘properties’ /khûapkhum/ ‘control’ /thǎankhǒmuun/ ‘data base’ /ʔànaamay/ ‘hygiene’ /náksàdɛɛŋ/ ‘actor’ /phǒtcháanaakaan/ ‘nutrition’ /phátthánaa/ ‘develop’ /canyaaban/ ‘ethics’ /kaanriankaansǒn/ ‘schooling’ /paanklaan/ ‘average’ /sǒəmsâan/ ‘enhance’ /náam/ ‘water’ /kaansùksǎa/ ‘education’ /khunnátham/ ‘virtue’ /**phǒnkràthóp**/ ‘effect’ /khunnáláksàná/ ‘characteristic’ /lákpràkan/ ‘security’ /piam/ ‘full of’ /ràksǎa/ ‘cure’ /brɛen/ ‘brand’ /**tám**/ ‘low’ /sǎn/ ‘select’ /phianphǒw/ ‘adequate’ 2544

The collocates of *quality* ranked by Log Ratio score are as follows:

merchantable bs5750 tqm 5750 assurance redeeming workmanship watchword craftsmanship indefinable qa quantity reliability dependable birkett intrinsic improve teamwork knitwear groundwater **deterioration** bs endearing affordable accreditation timeless durability improving **elusive** **inferior** **deteriorating** accessibility improvement nutritional improved healthcare admirable cleanliness aesthetic appropriateness **seductive** fleece certification acoustic fitness reproduction rarity enhance **deduct** stringent audit possess exceptional excellence productivity **deteriorate** high harmonic leadership bthe superior charismatic **poor** renowned superb proven masculine artistic innate photographic audio product inherent threshold effectiveness recording vastly output printing standard breadth rigorous reputation **impair** teaching consistency magical beef outstanding consistently management architectural monitor tabloid carpet varying impressed strive specification indicator life riding commitment equate service warranty drinking control air monitoring efficiency printed admire circle guarantee assess packaging excellent consistent availability delivery desirable • acceptable

6.5.1.48 /kitcàkaan/ vs. *business*

The collocates of /kitcàkaan/ ranked by Log Ratio score are as follows:

/rûamkháa/ ‘joint venture’ /pràmoŋ/ ‘fishery’ /thoorákhámánaakhom/ ‘telecommunication’ /phaanítcháyákam/ ‘commerce’ /khoonom/ ‘dairy cow’ /kràcaaysǎn/ ‘broadcast’ /khûap/ ‘combine’ /damnəən/ ‘manage’ /kamkàpduulɛɛ/ ‘supervise’ /pràkǒp/ ‘do’ /roonŋaan/ ‘factory’ /sǎathaarànuupàphǒk/ ‘public utility’ /rǒw fǒw mǒw/ (acronym for ‘Mass Rapid Transit Authority of Thailand’) /ʔoon/ ‘transfer’ /thálee/ ‘sea’ /camphûak/ ‘type’ /mǎanrêe/ ‘mine’ /lûuksǎa/ ‘boy scout’ /**sêeksɛɛŋ**/ ‘interfere’ /kǒw mǒw thǒw/ (acronym for ‘commission’) /kháakhǎay/ ‘trade’

/naaycâaŋ/ ‘employer’ /**sópsaw**/ ‘dull’ /**máhǎ́rásòp**/ ‘entertainment’ /**khánákammaathíkaan**/ ‘commission’ /lúukcâaŋ/ ‘employee’ /dǎənrɯa/ ‘sail’ 58 /**khúmkehrǎəŋ**/ ‘protect’ /rátwisǎahàkit/ ‘state-owned enterprise’ /**pràkanphay**/ ‘insurance’ /kiawnhâaŋ/ ‘relevant’ /wisǎaman/ ‘uncommon’ /**lómłǎək**/ ‘cancel’ /kit/ ‘business’ /pìtrooliam/ ‘petroleum’ /ʔànúkammaakaan/ ‘subcommittee’ /sǎmpàthaaŋ/ ‘concession’ /khǎəŋwát/ /bandaa/ ‘all’ /**kháapràweenii**/ ‘prostitute’ /níkhomʔúsǎahàkam/ ‘industrial estate’ /lǎək/ ‘stop’ /ʔànúyâat/ ‘allow’ /bâanmuaŋ/ ‘country’ /sǎpsǐn/ ‘wealth’ /praysàni/ ‘post’ /**phúuduulɛɛ**/ ‘supervisor’ /**nǐsǐn**/ ‘debt’ /cǎwkhǎəŋ/ ‘owner’ /**lómłálaay**/ ‘bankrupt’ /rɛŋŋaaŋ/ ‘labourer’ /**lonthun**/ ‘invest’ /**yûŋkiaw**/ ‘meddle’ /sǎthâaŋ/ ‘place’ /pràkǎəpkaan/ ‘do (a business)’ /**kràthuan**/ ‘affect’ /fayfáa/ ‘electricity’ /nítibùkkhon/ ‘juristic person’ /sǐnsáp/ ‘asset’ /rûŋrɯaŋ/ ‘prosper’ /prɛrɯup/ ‘transform’ /**fúunfuu**/ ‘restore’ /dǎysǎ/ ‘gain and lose’ /khâathamniam/ ‘fee’ /wíttháyú/ ‘radio’ /phrǎráatchákrítisàdiikaa/ ‘royal decree’ /kǎə kǎə tǎə/ (acronym for ‘The Election Commission of Thailand’) /khúru/ ‘teacher’ /hũnsuàn/ ‘(business) partner’ /ŋóp-prámaan/ ‘budget’ /bǎəríhâaŋ/ ‘administer’ /mǎəŋ/ ‘mine’ /nâathii/ ‘duty’ /**khàatthun**/ ‘lose one’s capital’ /sǎhàkǎəŋ/ ‘cooperative’ /**khàyǎay**/ ‘expand’ /pàamáy/ ‘forest’ /**sùupthǎət**/ ‘carry on’ /**phùukkhàat**/ ‘monopolise’ /kháa/ ‘trade’ /pràsǎŋ/ ‘wish’ /**khúkkhák**/ ‘vigorous’ /khǎənsǎŋ/ ‘transport’ /**kamray**/ ‘profit’ /faam/ ‘farm’ /rápcâaŋ/ ‘work for hire’

The collocates of *business* ranked by Log Ratio score are as follows:

owner-managed transact non-contentious acumen **non-core** gribben
 unincorporated **unfinished** re-engineering bradstreet mba blue-chip
 medium-sized over-the-counter self-employment thriving regulated
 courtauld refining booming start-up mail-order haulage reinsurance
 family-run harvard broking automotive **loss-making** franchisee profitable
 bes millar lucrative btec tailor as/400 resale minicomputer solicit
flourishing **tout discontinued** telecoms acrylic day-to-day 11 9000 multi-
 million goodwill mainframe **risky ailing** ici oem tycoon bae hairdressing
underwriting conduct core accountancy investment leasing going
laborious banking dealing brisk messy prosper consultancy self-employed
 aea sized high-end mid-range secretarial commerce shrewd venture
 management ethics **tedious** stationery state-owned expand profitability
 taxable consulting retail reorganise contracting sponsorship **tricky**
 economics premise printing exempt authorisation partnership acquaintance
 enterprise wholesale sparc barclays conglomerate relocate biography
catering entrepreneurial pharmaceutical publishing aerospace vendor tec
 coating franchise **bust** expanding restructure at&t **time-consuming**
 warner competitiveness accounting dictionary technician cbi hp warwick
 corporate semiconductor workstation lawful asset promotional purchaser
viable tenancy traveller **costly**

6.5.2 Methodological issues

One methodological issue I encountered in generating the data above was that it was sometimes difficult to classify collocates as positive, negative, or neutral based on the collocate lists alone, both with the Thai data and the English data. With the Thai data, it was sometimes necessary to examine the concordance lines where a collocate occurs. Out of the 3,531 collocate types examined for the 48 different nodes, there were 320 collocate types (approximately 9%) where I had to look at the concordance lines in order to provide a precise English translation and to accurately classify the collocate.

It was necessary to look at the concordance lines for two reasons. First, some collocates are erroneously tokenised. Tokenisation errors affecting the collocates can be broadly classified into two groups. Some collocates are erroneously split, whereas others are erroneously combined.

There are two major groups of collocates that tend to be erroneously split. The first group is proper nouns, usually either a person's name or a place name, whether Thai or foreign. There are several instances where a name is not tokenised as one single word. Instead it is split into two or more words. One or more of these erroneously-split words then becomes a collocate. For example, /yút/ appears as one of the collocates of /naayókrátthàmontrii/ 'prime minister'. This element is in fact part of /sùráyút/, a Thai former PM's name. In this case and in all other similar cases, I gave the translation as (part of a person's name) in 6.5.1 above. In cases where an erroneously-split word is part of a place name, I gave the translation as (part of a place name). However, in cases where the proper noun refers to a very common concept, its English translation was provided. One example is /máaksít/ 'Marxism'. /máaksít/ is erroneously split into /máak/ and /sít/.

/máak/ appears as a collocate of /pràwàtsàat/ ‘history’ and /nɛɛwkhít/ ‘idea’, whereas /sít/ appears as a collocate of /nɛɛwkhít/. In both cases, the translation ‘Marxism’ was given.

The second group of words found to have been incorrectly split is English loanwords. For example, /hanlǒo/ ‘hello’ is split into /han/ and /lǒo/. Both /han/ and /lǒo/ appear as collocates of /thoorásàp/ ‘telephone’. In these cases, ‘hello’ was given as a translation for both /han/ and /lǒo/, and both were classified according to the meaning of ‘hello’, which is neutral. Other such cases were treated in the same way.

It is quite understandable why proper nouns and loanwords are likely to be inaccurately tokenised. The corpus is automatically tokenised, and it seems that it is hard for the machine to recognise all the proper nouns and loanwords. This is probably because these proper nouns and loanwords are not on the list of known words that the machine tokeniser is using.

There are also cases where words other than a proper noun or a loanword are erroneously split. One example is /trítsàdii/. /trítsàdii/ is a simplex word that means ‘theory’. It is erroneously split into /trítsà/ (a non-word) and /dii/ (which can be a word meaning ‘good’, but is not here), both of which occur as collocates of /phúunthään/ ‘foundation’. In this case, I gave a translation of both collocates as ‘theory’ and classified them according to the meaning of ‘theory’. Other similar cases where a simplex word is erroneously split were treated the same way.

The above are cases where I was certain that there is an error in tokenisation. However, there are other cases where I would rather not state confidently that there is an error in tokenisation. One example is /níphon/. /níphon/ means ‘writing’. It occurs as a collocate of /pràwàtsàat/ ‘history’. Examining the concordance lines of /níphon/ as a

collocate of /pràwàtsàat/, I found that /níphon/ is in fact part of three different words: /sǎará níphon/ ‘dissertation’, /wítháyaa níphon/ ‘thesis’, and /ŋaan níphon/ ‘piece of writing’. These words are complex words consisting of two morphemes, one of which is /níphon/. On the basis that (a) /níphon/ is in itself a meaningful unit, and (b) throughout the analysis, I followed the tokenisation in the corpus, except only when I was confident of there being an error, I considered /níphon/, rather than /sǎará níphon/, etc., as a collocate of /pràwàtsàat/, and classified the word as neutral accordingly. Other similar cases were treated in the same way.

Whether the case of /níphon/ is an instance of a tokenisation error is debatable. Some might argue that it is erroneously tokenised, in that it is part of a complex word, which should be treated as a single unit and not divided. Others might object that although it is part of a complex word, it is on its own a meaningful unit, and should therefore be tokenised as a single unit. In fact, the tokenisation of such complex words in the corpus seems in general to be inconsistent. Whereas some complex words are tokenised as a single unit, others are separated into different units. However, as tokenisation is not the focus of my study, I mostly followed the tokenisation in the corpus, (unless I was confident that the given collocate is a result of tokenisation errors, as previously mentioned). Furthermore, whether /níphon/ or /sǎará níphon/, etc., is counted as the collocate of the node does not affect the result of the analysis in this case, as all the words in question are neutral in meaning.

However, there are also cases where the tokenisation of particular complex words adopted in the corpus does seem to affect the result of the analysis. One instance is /boo/. /boo/ alone means ‘bow (as in bow tie)’. It is a collocate of /phǒnŋaan/ ‘achievement’.

Examining the concordance lines of /boo/ as a collocate of /phǒnŋaan/, I found that /boo/ is split from /boo dɛɛŋ/, a compound noun meaning ‘masterpiece’. /boo dɛɛŋ/ *also* in fact occurs as a collocate of /phǒnŋaan/. Since like /níphon/, /boo/ is in itself a meaningful unit, to be consistent, I counted /boo/, rather than /boodɛɛŋ/, as a collocate of /phǒnŋaan/. However, in contrast to the case with /níphon/, /boo/ and /boo dɛɛŋ/ have a different evaluative meaning. While /boo/ is neutral, /boo dɛɛŋ/ is positive. Thus, in the analysis of /phǒnŋaan/, the number of neutral collocates, rather than the number of positive collocates, increases when /boo/ is considered. This affects the total numbers of positive and neutral collocates, which might in turn sway the semantic prosody of the node. Fortunately, cases like this one are rare in the analysis; there were actually only four cases including this one.

There are a relatively few cases where the collocates are erroneously combined. One example is /tǒŋhǎa/. /tǒŋhǎa/ appears as a collocate of /khamtǒp/ ‘answer’. Here, /tǒŋ/ is a modal auxiliary ‘must’ that is followed by a verb /hǎa/ which means ‘find out’. /tǒŋhǎa/ is therefore not a single word, but two separate words. /pràkǒpkaan/ is another example. /pràkǒpkaan/ in itself means ‘do (a business)’. It appears as a collocate of /kitcàkaan/ ‘business’. Nevertheless, there are instances of co-occurrence of /pràkǒpkaan/ and other nodes where /pràkǒpkaan/ results from incorrect tokenisation of the (distinct) word /pràkǒp/ ‘supplement’, and /kaan/, a nominalising particle. In these two cases and all other similar cases, I left out the translation and classified the collocates as neutral. This is because it was incorrect data, and classifying the collocates as either positive or negative would sway the semantic prosody of the node. By contrast, labelling

them as neutral leaves the semantic prosody, as determined from the other collocates, unchanged.

Second, it was sometimes hard to figure out the meaning of the collocate and classify it into the accurate meaning category out of context. There were cases where I simply could not figure out the meaning and translation of the collocate out of context. One example is /phlík/. /phlík/ is a collocate of /sàthäänákaan/ ‘situation’. Out of context, I could not figure out the exact meaning of /phlík/ apart from its general meaning of ‘turn over’. However, its meaning became more obvious when I looked at the word in context. In context, /phlík/ means ‘change’, as in ‘change the situation’.

There are several cases where the collocate has more than one meaning, and the intended meaning can only be figured out from context. For example, /pràkòp/¹² collocates with /rátthàthammánuun/ ‘constitution’, /ʔeekkàsään/ ‘document’, and /kitcàkaan/ ‘business’. /pràkòp/ has several meanings, such as ‘supplement’, ‘do’, and ‘assemble’. Looking at the context allowed me to assign the most precise translation to each collocate. Specifically, when /pràkòp/ co-occurs with /rátthàthammánuun/ and /ʔeekkàsään/, it means ‘supplement’. On the other hand, when it appears with /kitcàkaan/, it means ‘do (a business)’. Both meanings led, however, to a classification as neutral. In this case, the different meanings of the collocate do not seem to affect the analysis, as they all fall into the same meaning category.

There are also cases where the different meanings of the collocate fall into the different meaning categories. One example is /bòtrian/. /bòtrian/ co-occurs with /pràwàtsàat/ ‘history’, /pràsòpkaan/ ‘experience’, and /nákrian/ ‘student’. It has two

¹² /pràkòp/ is a separate collocate from /pràkòpkaan/, which I discussed above. They both occur as collocates of /kitcàkaan/.

meanings. One is ‘unit of instruction’. The other is ‘something that you learn from life, such as a bad experience that teaches you’. Both meanings can be translated into ‘lesson’ in English. In context, however, /bòtrian/ means ‘bad experience’ when it co-occurs with /pràwàtsàat/ and /pràsòpkaan/, but it means ‘unit of instruction’ when it appears with /nákrian/. It is therefore classified as negative under /pràwàtsàat/ and /pràsòpkaan/, but neutral under /nákrian/. This accurate classification would not have been possible without looking at the concordance lines.

We have seen that, due to the reasons outlined above, examining concordance lines is sometimes necessary even within the polarity-oriented approach. I also found a case where looking at the concordance lines of a collocate enabled me to discover not only the intended meaning and evaluation of a collocate, but also something else. /tòkpen/ ‘become’ collocates with two of the nodes in this study: /pâwmăay/ ‘target’ and /sápsîn/ ‘wealth’. /tòkpen/ in itself is a neutral word, and I classified it as neutral in my analysis. What was interesting was when I looked at the concordance lines of /tòkpen/ as a collocate of /pâwmăay/, I found that in 25 out of 26 instances, /tòkpen/ is used in the pattern /tòkpen pâwmăay/ ‘become the target’. Of these 25 instances, there are 18 where /tòkpen pâwmăay/ occurs in a negative environment, such as /tòkpen pâwmăay khǒṅ mitchăachîp/ ‘become the target of a gang of thieves’. It might thus be argued that, although /tòkpen/ in itself is neutral, when it co-occurs with /pâwmăay/, their co-occurrence tends to form an extended unit of meaning with a negative semantic prosody that belongs to the unit as a whole. We could also say, from the lexical priming perspective – which for Partington plays a crucial role in the discussion of semantic prosody (see section 2.4) – that /tòkpen pâwmăay/ is a nesting unit. As a nesting unit, it

can have its own priming(s), which in this case, is a particular semantic prosody, that is, its tendency to be used in negative contexts. This priming belongs to the nesting unit as a whole, and does not apply individually to either /tòkpen/ or /pâwmăay/.

One might then argue that, as /tòkpen pâwmăay/ as an extended unit of meaning has a negative semantic prosody overall, /tòkpen/ should be classified as a negative collocate here. However, I did not follow this approach: I kept /tòkpen/ classified as neutral. Within the polarity-oriented approach, we only look at the positive or negative meaning of individual collocates (see section 3.4.1). Looking at an extended unit of meaning in Sinclair's sense is the practice of the EUM-oriented approach. Thus, from the present perspective, the fact that the unit /tòkpen pâwmăay/ may seem to have a negative semantic prosody does not make /tòkpen/ a negative word in itself or contribute to analysis of /pâwmăay/ as having a negative semantic prosody. In fact, as will be discussed later, /pâwmăay/ is found to have a positive semantic prosody.

As previously mentioned, it was also sometimes difficult to classify the English collocates as positive, negative, or neutral. As a non-native speaker of English, there are cases where I was not sure whether certain collocates had positive, negative, or neutral evaluative meaning. Approximately 20% of collocate types were of this sort. I dealt with these cases by seeking an opinion from a native speaker of English who has a background in linguistics.

In discussion with this native speaker, I found that my informant also sometimes had difficulty classifying the collocates into the meaning categories. This is primarily because, as my informant said during the exercise, the evaluative meaning of the collocates sometimes depends on the context in which they occur, or on what node the

collocates co-occur with. This is the same experience I had had when analysing the Thai data. One example is *knockout*. It occurs as the collocate of *village*. In this case, my informant was not sure whether the collocate was positive or negative, reasoning as follows: *knockout* has two completely different senses. It can mean ‘an attractive woman’, which is positive, or ‘a hit that knocks someone down’, which is negative. In this situation, we needed to look at the concordance lines to figure out what sense is being used in context in order to classify the collocate accurately. *Knockout* is quite a difficult case, because it was necessary to look at the concordance. There are easier cases where simply informing my informant what node the collocates co-occur with enabled them to classify the collocates. One example is *shady*. My informant was at first torn between positive and negative. However, after they were informed that *shady* is the collocate of *place*, they were certain that *shady* in this context has a positive evaluative meaning. There are also cases where the evaluative meaning of the collocate is contingent upon the collocate’s part of speech. One example is *loot*. As a noun, *loot* is positively evaluated, but as a verb, it is negatively evaluated. Only looking at the collocate’s concordance lines made an accurate classification possible.

We have therefore seen that with both Thai and English data, we cannot always straightforwardly classify collocates as positive, negative, or neutral simply based on the collocate lists. It is often necessary to look at the context, for various reasons outlined above.

The difficulty in classifying the collocates as positive, negative, or neutral, both in Thai data and in English data, also lies in the fact that some collocates are technical terms from specialised domains, such as law, politics, and finance. Most of these collocates are

difficult to classify, because they are little used outside their domains. It was thus sometimes difficult for me and my informant to label these collocates. Moreover, some collocates could not be analysed at all, since they are part of a fixed expression. Two examples from the English data are *prima* and *facie*, both of which are part of the legal jargon fixed phrase *prima facie*. Of the 48 Thai nodes, it might be argued that 9 are from specialised domains: /rátthàthammánuun/ ‘constitution’, /naayókrátthàmontrii/ ‘prime minister’, /khamràj/ ‘order’, /lákthǎan/ ‘evidence’, /sápsǐn/ ‘property’, and /hóp-pràmaan/ ‘budget’ from law and politics, and /ʔèkkàchon/ ‘individual’, /ʔonkoon/ ‘company’, and /kitcàkaan/ ‘business’ from finance. It might be argued that these nodes, and their English translation-equivalents, are not suitable for this kind of analysis, because the collocates of specialised jargon are likely to be technical and thus hard for a non-specialist to label as positive, negative, or neutral. On the other hand, these are hard cases. So we can assume that if the analysis works with these difficult cases, it should work at least as well with easy cases, such as non-technical terms. In fact, since the analysis includes only nine technical nodes, the majority of the nodes under study are non-technical. Therefore, it is worth investigating these technical nodes along with the more usual non-technical nodes.

There are a few cases where a collocate in itself has a clear evaluative meaning out of context, but the evaluative meaning depends on whom you ask. One example from the Thai data is /sǎŋkhomníyom/ ‘socialism’. For people with socialist politics, /sǎŋkhomníyom/ is very likely to be positive. Those who do not support this political system are likely to dislike “socialism” and thus negatively evaluate the word. A similar example from the English data is *Conservative*. People who support the Conservative Party might label the collocate as positive, whereas those who oppose this political party

might label it as negative. In these situations, I labelled the collocates as neutral. This is not because the collocates in themselves are neutral, as previously discussed. But labelling the collocates as either positive or negative would affect the quantitative result of the analysis. It would sway the quantitative result for this node in either a positive or a negative direction, when it should not. Classifying the collocates as neutral, on the other hand, will effectively exclude these collocates from pushing the quantitative result in either a positive or a negative direction. That is, the quantitative result for node will be decided on its other collocates. The presence of collocates of this sort will thus weaken the strength of the semantic prosody, but not sway it the wrong way.

There are some spelling issues that need to be mentioned. In the Thai data, there are a few cases where a similar collocate appears twice. This is due to spelling mistakes. That is, one collocate is spelt correctly, whereas the other is misspelt, and both appear in the collocate list. In this situation, only one collocate was given in the data.

The Thai alphabet has its own numerals. There are a few cases where a Thai numeral and the corresponding Arabic numeral both appear as collocates. In this situation, only the Arabic numeral was given. In cases where only a Thai numeral appears in the collocate list, the equivalent Arabic numeral was given instead of the Thai numeral as a transliteration.

In the English data, there are many cases where a similar collocate appears twice. This is due to differences in spelling. In some cases, one collocate has a British spelling, and the other has an American spelling. In other cases, one collocate is hyphenated, and the other is not. In this situation, only the collocate that appears first in the list, that is, the collocate that has a higher Log Ratio score, was presented.

There are also errors in the BNC's lemmatisation. There are a few cases where the lemmatiser has failed to recognise the correct lemma and has presented an incorrect guess. In this situation, I corrected it. In cases where lemmatisation errors have resulted in forms of the same lemma being listed under two lemmas, I merged the two together.

6.5.3 Semantic prosody of each Thai word and its English translation-equivalent

Table 6.2 summarises the number of positive, negative, and neutral collocate types, the semantic prosody, and the prosodic strength of each Thai word and its English translation-equivalent. The prosodic strength is presented only for the cases where both words in a translation pair have the same positive or negative semantic prosody. See section 3.4.1.1 for the criteria for assigning a positive, negative, and neutral semantic prosody to a word.

	Thai word/English translation-equivalent	Collocate types/percentage of types				Semantic prosody	Prosodic strength
		Positive	Negative	Neutral	Total		
1.	/phǒnŋaan/ achievement	23/0.33	1/0.01	45/0.66	69	Positive	23
		38/0.52	2/0.03	33/0.45	73	Positive	19
2.	/thǒŋthìn/ district	16/0.24	1/0.01	50/0.75	67	Neutral	-
		4/0.03	5/0.03	135/0.94	144	Neutral	-
3.	/pràwàtsàat/ history	10/0.13	3/0.04	65/0.83	78	Neutral	-
		17/0.12	10/0.07	1120.81	139	Neutral	-
4.	/thoorásàp/ telephone	2/0.02	8/0.10	74/0.88	84	Neutral	-
		6/0.05	7/0.06	100/0.89	113	Neutral	-
5.	/phûukhon/ people	9/0.13	28/0.41	32/0.46	69	Negative	3.1
		14/0.08	46/0.27	110/0.65	170	Negative	3.3
6.	/phâappháyon/ film	7/0.09	1/0.01	73/0.90	81	Neutral	-
		15/0.06	14/0.05	237/0.89	266	Neutral	-
7.	/pràsithíphâap/ efficiency	28/0.53	5/0.10	19/0.37	52	Positive	5.6
		39/0.44	4/0.04	46/0.52	89	Positive	9.8
8.	/ŋûankhăy/ condition	5/0.13	4/0.10	31/0.77	40	Neutral	-
		36/0.19	79/0.43	71/0.38	186	Neutral	-

Table 6.2 The number of positive, negative, and neutral collocates, semantic prosody, and prosodic strength of each Thai word and its English translation-equivalent

	Thai word/English translation-equivalent	Collocate types/percentage of types				Semantic prosody	Prosodic strength
		Positive	Negative	Neutral	Total		
9.	/phûuyày/ adult	15/0.28	3/0.06	35/0.66	53	Positive	-
		14/0.18	9/0.11	56/0.71	79	Neutral	-
10.	/sùkkhàphâap/ health	32/0.48	11/0.16	24/0.36	67	Neutral	-
		36/0.19	32/0.17	121/0.64	189	Neutral	-
11.	/rátthàthammánuun/ constitution	8/0.09	7/0.08	70/0.83	85	Neutral	-
		28/0.25	10/0.09	76/0.66	114	Neutral	-
12.	/pâwmăay/ target	13/0.34	3/0.08	22/0.58	38	Positive	-
		25/0.26	10/0.10	62/0.64	97	Neutral	-
13.	/chaaynùm/ young man	14/0.12	20/0.17	86/0.71	120	Neutral	-
		15/0.56	2/0.07	10/0.37	27	Positive	-
14.	/sîṅwêetlôom/ environment	27/0.35	11/0.14	39/0.51	77	Neutral	-
		38/0.17	25/0.11	165/0.72	228	Neutral	-
15.	/pràsòpkaan/ experience	19/0.48	5/0.12	16/0.40	40	Positive	-
		48/0.37	27/0.21	54/0.42	129	Neutral	-
16.	/khwaamtôṅkaan/ desire	16/0.37	2/0.05	25/0.58	43	Positive	-
		25/0.29	27/0.32	33/0.39	85	Neutral	-
17.	/tàwantòk/ west	12/0.13	8/0.09	71/0.78	91	Neutral	-
		5/0.02	3/0.01	213/0.97	221	Neutral	-
18.	/ʔitthíphon/ influence	8/0.12	6/0.09	51/0.79	65	Neutral	-
		29/0.31	16/0.17	48/0.52	93	Neutral	-
19.	/chûamooṅ/ hour	2/0.03	1/0.01	66/0.96	69	Neutral	-
		9/0.06	16/0.11	127/0.83	152	Neutral	-
20.	/phúunthāan/ foundation	16/0.29	1/0.02	38/0.69	55	Positive	-
		15/0.18	4/0.05	63/0.77	82	Neutral	-
21.	/sàthāanākaan/ situation	15/0.24	28/0.39	28/0.37	71	Neutral	-
		25/0.19	68/0.52	38/0.29	131	Neutral	-
22.	/khamtòṅp/ answer	4/0.12	6/0.18	23/0.70	33	Neutral	-
		17/0.35	7/0.14	25/0.51	49	Neutral	-
23.	/raayláʔiat/ detail	7/0.13	1/0.02	45/0.85	53	Neutral	-
		17/0.22	11/0.14	49/0.64	77	Neutral	-
24.	/phômêe/ parents	15/0.31	11/0.22	23/0.47	49	Neutral	-
		12/0.16	18/0.25	43/0.59	73	Neutral	-
25.	/nēwkhít/ idea	9/0.13	4/0.06	56/0.81	69	Neutral	-
		24/0.31	26/0.34	27/0.35	77	Neutral	-
26.	/ʔèekkàchon/ individual	17/0.24	1/0.01	54/0.75	72	Neutral	-
		27/0.40	13/0.20	27/0.40	67	Neutral	-
27.	/naayókrátthàmontri/ prime minister	8/0.04	3/0.02	169/0.94	180	Neutral	-
		6/0.02	3/0.01	390/0.97	399	Neutral	-
28.	/ʔèekkàsāan/ document	10/0.10	7/0.07	82/0.83	99	Neutral	-
		29/0.22	11/0.08	92/0.70	132	Neutral	-

Table 6.2 The number of positive, negative, and neutral collocates, semantic prosody, and prosodic strength of each Thai word and its English translation-equivalent

	Thai word/English translation-equivalent	Collocate types/percentage of types				Semantic prosody	Prosodic strength
		Positive	Negative	Neutral	Total		
29.	/nɛwʔhaan/ way	23/0.47	1/0.02	25/0.51	49	Positive	-
		16/0.29	13/0.23	27/0.48	56	Neutral	-
30.	/muangʔhaya/ Thailand	8/0.36	1/0.05	13/0.59	22	Positive	-
		3/0.07	1/0.02	41/0.91	45	Neutral	-
31.	/lákʔhãan/ evidence	18/0.18	13/0.13	68/0.69	99	Neutral	-
		41/0.27	39/0.25	73/0.48	153	Neutral	-
32.	/sápsǐn/ property	32/0.32	26/0.26	42/0.42	100	Neutral	-
		42/0.23	34/0.19	103/0.58	179	Neutral	-
33.	/kʔhamsàŋ/ order	12/0.14	20/0.23	56/0.63	88	Neutral	-
		36/0.24	29/0.19	84/0.57	149	Neutral	-
34.	/ŋópʔpʔraamaan/ budget	23/0.20	9/0.08	85/0.72	117	Neutral	-
		24/0.18	16/0.12	94/0.70	134	Neutral	-
35.	/náamman/ oil	13/0.08	13/0.08	147/0.84	173	Neutral	-
		31/0.10	34/0.11	252/0.79	317	Neutral	-
36.	/kʔhãarâatchákaan/ civil servant	18/0.17	4/0.04	84/0.79	106	Neutral	-
		9/0.31	1/0.03	19/0.66	29	Positive	-
37.	/ʔoŋkɔɔn/ company	18/0.26	2/0.03	48/0.71	68	Neutral	-
		37/0.13	15/0.05	226/0.82	278	Neutral	-
38.	/cítcaj/ mind	52/0.46	39/0.34	23/0.20	114	Neutral	-
		22/0.29	29/0.38	26/0.33	77	Neutral	-
39.	/mãatʔtrãthãan/ standard	24/0.32	3/0.04	47/0.64	74	Positive	8
		80/0.43	17/0.09	89/0.48	186	Positive	4.7
40.	/nákrían/ student	11/0.18	1/0.02	48/0.80	60	Neutral	-
		34/0.24	8/0.05	102/0.71	144	Neutral	-
41.	/kʔhwaamsãamãat/ ability	26/0.46	10/0.18	21/0.36	57	Neutral	-
		62/0.50	16/0.13	45/0.37	123	Positive	-
42.	/mũubãan/ village	7/0.10	2/0.03	59/0.87	68	Neutral	-
		23/0.16	12/0.08	110/0.76	145	Neutral	-
43.	/sãthãanthĩi/ place	16/0.26	7/0.11	39/0.63	62	Neutral	-
		21/0.28	14/0.19	39/0.53	74	Neutral	-
44.	/chaawbãan/ villager	7/0.12	16/0.27	36/0.61	59	Neutral	-
		1/0.07	4/0.29	9/0.64	14	Negative	-
45.	/kítçákam/ activity	18/0.36	1/0.02	31/0.62	50	Positive	-
		68/0.33	41/0.20	98/0.47	207	Neutral	-
46.	/bòtbãat/ role	10/0.26	2/0.05	26/0.69	38	Positive	5
		60/0.40	14/0.09	77/0.51	151	Positive	4.3
47.	/khunnáphãap/ quality	28/0.40	8/0.11	35/0.49	71	Positive	3.5
		58/0.47	9/0.07	57/0.46	124	Positive	6.4
48.	/kítçákakaan/ business	16/0.19	10/0.11	61/0.70	87	Neutral	-
		39/0.27	14/0.10	89/0.63	142	Neutral	-

Table 6.2 The number of positive, negative, and neutral collocates, semantic prosody, and prosodic strength of each Thai word and its English translation-equivalent

6.6 Discussion

The analysis shows that of the 48 translation-equivalent pairs, 36 pairs display the same semantic prosody. However, of these 36 pairs, 30 have a neutral semantic prosody, that is, they do not show a particular tendency to co-occur with either positive or negative words. Five of the matching pairs have a positive semantic prosody; they are /phǒnɲaan/ ‘achievement’, /pràsithíphâap/ ‘efficiency’, /mâattràthǎan/ ‘standard’, /bòtbàat/ ‘role’ and /khunnáphâap/ ‘quality’. Only one matching pair shows a negative semantic prosody: /phûukhon/ ‘people’.

The finding that /phǒnɲaan/ ‘achievement’, /pràsithíphâap/ ‘efficiency’, /mâattràthǎan/ ‘standard’, and /khunnáphâap/ ‘quality’ show a positive semantic prosody in both languages may not come as a surprise. They all have positive basic meanings, so one would expect them to occur in positive environment. As Partington (2014: 283) argues, the evaluative meaning of positive or negative words is already apparent from their basic semantics, even without looking at their context. On the other hand, it is quite interesting that /bòtbàat/ ‘role’ and /phûukhon/ ‘people’ display respectively a positive semantic prosody and a negative semantic prosody. Unlike the previously discussed four pairs, the evaluation of /bòtbàat/ ‘role’ and /phûukhon/ ‘people’ is non-obvious. Without looking at the words’ collocates, one could hardly predict whether they would tend to occur in positive, negative, or neutral environment. This is probably the reason why most of the words whose semantic prosody has been studied in the literature are semantically neutral (Stewart 2010: 32). These words have been studied precisely because the part of their semantics that is immediately obvious is neutral, and the interest of scholars since Louw has been in identifying the non-neutral part of their meaning expressed through

patterns of co-occurrence. For this reason, one might argue that the analysis did not reveal much about /phǒnɲaan/ ‘achievement’, /pràsithíphâap/ ‘efficiency’, /mâattràthǎan/ ‘standard’, and /khunnáphâap/ ‘quality’ as it did with /bòtbàat/ ‘role’ and /phûukhon/ ‘people’. The finding that these positive words tend to occur in positive environments in both languages is perhaps not as interesting as the finding that identifies the intuitively neutral words as having a tendency to occur in positive or negative contexts, again across languages. That said, we also see many cases (among the previously mentioned 30 matching pairs) where words with neutral basic meanings do not show any particular tendency to co-occur with either positive or negative words.

One interesting point about the pair /phûukhon/ ‘people’ is that although both Thai and English words show a negative semantic prosody, they appear to have different semantic *preferences*. /phûukhon/ has semantic preferences for words expressing the qualities of being crowded and chaotic. They are expressed through collocates such as /phlúkphlâan/ ‘in disorder’, /khwàkkhwàw/ ‘helter-skelter’, /cɔɔcɛɛ/ ‘crowded and noisy’, /biàtsiat/ ‘congested’, /tɛ̀ɛktùn/ ‘panic’, /ʔonlàmaan/ ‘confusing’, /nɲaɲnêen/ ‘crowded’, and /muɲ/ ‘crowd around’. ‘People’, on the other hand, has a semantic preference for illness, which is absent in the collocate list of /phûukhon/. Some examples of relevant collocates are *hiv/aids, handicapped, deaf, disordered, hospitalize, disability, frail, and injure*. There are, however, overlapping semantic preferences for death and starvation: in Thai /ʔòtyàak/ ‘starve’ and /lóm/ ‘die, fall’, in English *killed, behead, kill, and starvation*. These semantic preferences are not, however, as obvious as the preferences for the qualities of being crowded and chaotic and for illness. (See section 6.5.1.5 for the full collocate lists of /phûukhon/ and ‘people’.)

There are 12 translation-equivalent pairs where the English and Thai words have different semantic prosodies. In all these cases, one word in the pair has a positive or negative semantic prosody, and the other has a neutral semantic prosody. There are no cases of a word in one language having a positive semantic prosody, and the translation-equivalent having a negative semantic prosody. As with the pairs /bòtbàat/ ‘role’ and /phûukhon/ ‘people’, what we *do* see are some cases where words with neutral basic meanings have a tendency to occur in positive or negative contexts. However, in these 12 pairs, the semantic prosody is different between the two languages.

We have so far seen that most of the translation-equivalent pairs under study have the same semantic prosody. Among the cases where the words in the pair have different semantic prosodies, there is no case where one word has a positive prosody, and the other has a negative prosody. This shows that there seems to be a relationship between semantic prosody in Thai and semantic prosody in English. Otherwise, we would expect by chance alone at least some cases where one word in the translation-equivalent pair has a positive semantic prosody and the other has a negative semantic prosody. In fact, most studies that feature a comparison of semantic prosody in English and another language have shown similar results. That is, the English words and their translation-equivalents in another language display the same semantic prosody, although in some cases they do not have the same prosodic *strength*, a topic I will return to later. The literature does identify a very few cases, however, where English words and their equivalents show the opposite semantic prosody. (See section 2.7 for a review of studies of semantic prosody from a cross-linguistic perspective.)

One possible reason why most of the translation-equivalent pairs under study share the same semantic prosody could be simply that they are translation-equivalents of each other, so they basically share the same core semantics and are thus likely to occur in similar evaluative contexts. That said, there are also some cases where the English translation-equivalent has senses that do not exist for its Thai counterpart, and vice versa. For instance, according to the *Macmillan Dictionary* of English, ‘condition’, the translation-equivalent of /n̄h̄ankhǎy/, has three senses that do not exist for /n̄h̄ankhǎy/, which only means ‘something that must be true or done before another thing can happen’. There are seven more pairs in my set of nodes where one word obviously has senses that the other does not have. These pairs are /phâappháyon/ ‘film’, /siŋwêetlóm/ ‘environment’, /n̄ewthaaŋ/ ‘way’, /sápsǐn/ ‘property’, /khamràŋ/ ‘order’, /ʔoŋkɔɔn/ ‘company’, and /kitcàkaan/ ‘business’. However, despite the differences in the range of senses, seven of these eight pairs have the same neutral semantic prosody. The only pair that has different semantic prosodies is /n̄ewthaaŋ/ ‘way’, where /n̄ewthaaŋ/ has a clear positive semantic prosody, while ‘way’ has a neutral semantic prosody (albeit with more positive collocates than the negative ones).

In cases where the words in a pair display different semantic prosodies, it seems clear that the words in that pair can be equivalent only in terms of their basic semantics. In these cases, it could therefore be argued that the words are not direct equivalents, as they differ in terms of semantic prosody.

In cases where the words in a pair share a positive or negative semantic prosody, they may differ in terms of semantic preference, as we have seen with the pair /phûukhon/ ‘people’. They may also differ in terms of prosodic strength (see section 6.3),

or the degree of semantic prosody, in that one word may have a stronger prosody than the other. One example from the data is the pair /khunnáphâap/ ‘quality’. Both /khunnáphâap/ and ‘quality’ have a positive semantic prosody. However, the positive prosodic strength of ‘quality’ is higher than that of /khunnáphâap/ (quantified as 6.4 vs. 3.5). On these grounds, we could argue that /phûukhon/ and ‘people’ are not direct equivalents, because their semantic preference is different. We could similarly argue, following Ebeling (see section 6.3), that /khunnáphâap/ and ‘quality’ are not “real” equivalents, since their prosodic strength is different. In fact, of the six pairs where the words show the same positive or negative semantic prosody, there is not any pair where the prosodic strength of the two words is *exactly* the same. So if we follow Ebeling’s view, arguably, none of these pairs are “real” equivalents.

Among these six pairs, there are cases where the difference in prosodic strength is quite large, as well as cases where the difference is small. The difference in prosodic strength is quite large for the pairs /phõnṅaan/ ‘achievement’, /pràsithíphâap/ ‘efficiency’, /mâattráthăan/ ‘standard’, and /khunnáphâap/ ‘quality’. On the other hand, for the pairs /phûukhon/ ‘people’ and /bòtbàat/ ‘role’, the difference is very small. If we consider translation equivalence as a matter of degree, that is, in terms of Wei and Li’s degrees of equivalence, it follows that a large difference in prosodic strength implies a lower degree of equivalence, whereas a small difference in prosodic strength implies a higher degree of equivalence (see section 6.3).

We thus see that whereas Ebeling thinks in terms of whether the words in a translation pair are “real” equivalents or not, Wei and Li think in terms of degrees of equivalence. I would argue that it is more reasonable to think in terms of degrees of

equivalence rather than in all-or-nothing terms of whether the words are “real” equivalents or not. If we think in terms of “real” equivalents, any difference between the words in a pair, such as even a fairly modest difference in prosodic strength, will automatically rule the words out from being translation-equivalents. On the other hand, if we think in terms of degrees of equivalence, then rather than having an absolute criterion, the equivalence of a pair will be measured on a scale of prosodic strength. Thinking in terms of degrees of equivalence also makes better sense of my data. We have seen that of the six cases where the words in a translation pair show the same positive or negative semantic prosody, there is not any case in which the words in a pair show *exactly* the same prosodic strength. Thus, if I follow Ebeling, none of these pairs can be considered translation-equivalents. By contrast, if I follow Wei and Li, all these six pairs may be considered translation-equivalents, but with different degrees of equivalence.

Previously, I used the term “not direct” equivalents when discussing cases where the words in a translation pair have different semantic prosodies as well as the cases where the words in a pair have the same prosody, but different semantic preferences. Here, directness is a matter of degree of equivalence, rather than an absolute criterion. By describing the words as “direct” equivalents, I mean that the words have a relatively high degree of equivalence. On the other hand, by describing the words as “not direct” equivalents, I mean that the words have a relatively low degree of equivalence. They are translation-equivalents, but with obvious variation between them, i.e. in their semantic prosody and semantic preference in this context. Likewise here, rather than use the term “not real” equivalent (following Ebeling), I will also refer to cases where the words have the same semantic prosody, but different prosodic strength, as “not direct” equivalents.

Thus, following Wei and Li, I maintain that the words are still translation-equivalents, despite the differences in prosodic strength. Directness of equivalence in this sense exists on a scale. It is not all-or-nothing.

It was discussed earlier in section 6.5.2 that it was often difficult to classify both Thai and English collocates as positive, negative, or neutral simply based on the collocate lists. This difficulty highlights the problems with an operationalisation of the polarity-oriented approach based on looking at statistical collocates which assumes that we can classify collocates, just considered on their own, into evaluative meaning categories. In fact, we have seen many cases where the collocates, out of context, could not be straightforwardly classified as positive, negative, or neutral, as their evaluative meaning sometimes depends on what sense they are used in, on what context they occur in, or on what node they co-occur with. Just sticking to looking at the collocate lists was therefore not possible. Rather, we needed to look at the concordance lines. This unavoidable dependence on the concordance lines underlines the idea that applying the polarity-oriented approach via looking at statistical collocates alone is not enough. It always has to be tempered with the concordance analysis.

6.7 Concluding remarks

In this chapter, I have investigated similarities and differences in the semantic prosodies of translation-equivalents in Thai and English. From my results, I would argue that there is a relationship between semantic prosody in Thai and semantic prosody in English. The majority of the translation-equivalent pairs under study show the same semantic prosody. There are only 12 pairs where the words display different semantic

prosodies. However, in all these pairs, one word has a positive or negative prosody, and the other has a neutral prosody. There are no cases where one word has a positive semantic prosody, and the other has a negative semantic prosody.

In terms of cross-linguistic equivalence, I would argue that there do not *always* exist *direct* translation-equivalents, where the words in a translation pair have a high level of equivalence in all aspects. Instead, we have seen many cases where there are variations between the two words in terms of semantic prosody, semantic preference, and/or prosodic strength. Directness of equivalence seems always to be a matter of degree.

Some methodological issues arose during the process of doing the analysis. The polarity-oriented approach assumes that a collocate can simply be classified as positive, negative, or neutral out of context. We have seen, nevertheless, that with both Thai and English data it is in some cases difficult to classify the collocates into one of these meaning categories, as their evaluative meaning is contingent upon context. With the Thai data, difficulties can also arise from tokenisation errors. The fact that we need to depend on concordance lines underlines a shortfall of the polarity-oriented approach applied via looking at statistical collocates, in that it needs to be complemented by at least some use of the concordance analysis. These issues were not apparent when I examined the collocates of only three nodes in Chapter 4. They became obvious only when I looked at a large number of collocates across several dozen nodes.

Chapter 7 – Conclusion

7.1 Introduction

In this chapter, I summarise the findings of the study and lay out the answers to my research questions in section 7.2. I then discuss the study's limitations and the contributions to knowledge that it has made in section 7.3 and section 7.4 respectively. Finally, I will suggest some directions for possible future research to build on the outcomes of this study in section 7.5.

7.2 Findings of the study

This study aimed to carry out the groundwork of fundamental research into semantic prosody in Thai to set out the parameters for subsequent research in this area. This aim was accomplished by addressing three research questions. The approaches to semantic prosody that I adopted throughout my analyses were the polarity-oriented approach and the EUM-oriented approach. Through my review of the relevant literature, I established these to be the two primary approaches to semantic prosody that exist in the field, developed in the work respectively of Louw, Stubbs, and Partington and of Sinclair. My first research question (see section 1.2) was to consider the advantages and disadvantages of the two major approaches to the study of semantic prosody in Thai. Within the polarity-oriented approach, a semantic prosody is identified from collocates, and is restricted to the positive vs. negative opposition (see section 3.4.1.1). Within the EUM-oriented approach, by contrast, semantic prosody is identified from concordance lines, and can be any pragmatic function or meaning, rather than being confined to the

positive vs. negative opposition (see section 3.4.2). To address this research question, three Thai verbs, /kreeŋcay/ ‘considerate’, /kòŋhâykàət/ ‘cause’, and /chôp/ ‘like’, were selected, and their semantic prosody was examined using each of the two approaches.

The results obtained from the polarity-oriented approach (see section 4.4.1) show that /kòŋhâykàət/ and /chôp/ display a negative semantic prosody. /kreeŋcay/ does not display either a positive or negative semantic prosody, which is in effect a neutral semantic prosody, since it does not tend to co-occur particularly with either positive or negative words.

Using the EUM-oriented approach (see section 4.4.2), I identified some extended units of meaning for /kreeŋcay/ and /chôp/. These extended units of meaning have a clear pragmatic function or semantic prosody that is distinct from the literal meaning of the core of the unit, its specific colligations, and its semantic preferences. Unlike /kreeŋcay/ and /chôp/, /kòŋhâykàət/ does not seem to form part of any extended units of meaning within my data. Rather, it is *always* used as a straightforward verb, that is, a single-word unit of meaning.

These results allow us now to address the research question regarding the advantages and disadvantages of the major approaches to semantic prosody proposed in the literature for describing semantic prosody in Thai. The answer to this question would be that they are useful for different purposes. The polarity-oriented approach is useful when our purpose is to examine a tendency of a word to co-occur with positive or negative words. In particular, it reveals the hidden evaluative potential of a word whose evaluation is not immediately obvious in its core semantics. The EUM-oriented approach is, by contrast, suitable for the examination of an extended unit of meaning and its

pragmatic function in the Sinclairian sense. However, my subsequent analysis for the third research question (which I will summarise presently) led me to modify this conclusion somewhat; that analysis suggested that operationalising the polarity-oriented approach by looking at statistical collocates often needs to be complemented by at least some application of the kind of concordance analysis that I mainly utilised within the EUM-oriented approach.

The two approaches also have some advantages and disadvantages in terms of practicality. For example, we have seen that the practice of looking at individual collocates alone is not always sufficient to assess semantic prosody (see section 4.4.1.4). This is the disadvantage of the polarity-oriented approach. However, the polarity-oriented approach uses all the corpus data and is less time-consuming than the EUM-oriented approach, which uses much less data and is much more laborious. That said, the EUM-oriented approach produces the results that cannot be obtained from the polarity-oriented approach.

My second research question (see section 1.2) was what variation in semantic prosodies across genres can be identified for Thai words. The EUM-oriented approach was adopted in this study, since my purpose was to investigate whether the same word would participate in the same or different extended units of meaning, and correspondingly have the same or different pragmatic functions or semantic prosodies, across genres.

To address this research question, 19 verbs were selected, and their concordance lines were investigated in four different genres: *academic*, *fiction*, *newspaper*, and *non-academic* (see section 5.3 and section 5.4). The results of this analysis reveal that,

similarly to the results obtained from the concordance analysis in Chapter 4, each of the verbs under study, except /hǎn/, is mostly used as a single-word unit of meaning. That is, they are used independently, expressing no more than their meaning considered in isolation. That said, when used this way, they may still have collocations and/or colligations, and do still require the general colligations and semantic preferences characteristic of the broad class of verb to which they belong. This in turn creates specific patterns around the verbs. These patterns correspond to those described in Hunston and Francis' Pattern Grammar. In fact, the idea of a single-word unit of meaning that I call on here is also arguably congruent with Pattern Grammar (see argument in section 5.6.1).

Even though almost all the verbs under study are most frequently used as a single-word unit of meaning, some do also form part of an extended unit of meaning that has a clear semantic prosody. Of the 19 verbs, 14 appear to form part of one or more extended units of meaning. /càtkaan/, for example, forms part of the core of two different extended units of meaning, each of which has its own separate semantic prosody (see section 5.5.15).

Some of the extended unit of meaning that I identified in this analysis as well as in the analysis in Chapter 4 are susceptible to certain criticisms. That is, it might be argued, especially by scholars operating within the polarity-oriented approach, that some of the semantic prosodies are unusual, since they do not incorporate expressions of positive or negative evaluation, and moreover that some of them could be seen as simply a paraphrase of the core of the unit. To address this point, I have established that within the Sinclairian approach, semantic prosody can be any pragmatic meaning and is not limited to expressions of evaluation. I have also established that my analysis is less

vulnerable to the criticism of paraphrasing than those of some other scholars. That said, given that scholars working from the polarity-oriented perspective might still find my analysis problematic, I have in addition stepped out of the EUM-oriented approach, and attempted to characterise the polarity of evaluation of the semantic prosodies that I identified (see section 5.6.3).

We have seen variation across genres both when the verb is used as a single-word unit of meaning and when it forms part of an extended unit of meaning (see section 5.6.4 and section 5.6.5). Generally, variation exists in terms of frequency of occurrence of different patterns or of different extended units of meaning with these verbs across the four genres. Specifically, the level of variation in the use of the extended units of meaning, which also indicates variation in semantic prosodies, is considerable with some extended units of meaning, but limited with others. We have seen that while some extended units of meaning are genre-specific, in that they appear in only one, two, or three of the four genres, others appear across all four genres, but vary in their frequency of occurrence.

The variation in the use of extended units of meaning is in some cases explainable (see section 5.6.5). We have seen, for example, that the use of /yók tuayàaŋ chên/ is frequent in *academic*, but absent in *fiction*. This can be explained by the fact that the unit's pragmatic function of linking a statement of a phenomenon to an example of that phenomenon is rhetorically required in academic writing, but called for less in fiction. In fact, we have seen a contrast between *academic* and *fiction*, in that an extended unit that occurs frequently in one of these two genres is likely to be infrequent in the other genre. This is arguably due to the differences in the textual functions of the two genres. As

Biber *et al.* (1998: 149-152) argue, academic texts tend towards informational production, whereas fictional texts tend towards involved production, which is more associated with personal interactions.

Therefore, the answer to this research question would be that there does exist variation in semantic prosodies across genres. This variation is considerable with some extended units of meaning, but limited with others. Particularly, we have seen a contrast between academic writing and fiction.

The third and final research question (see section 1.2) aimed to examine to what extent the semantic prosodies of translation-equivalents in Thai and English are similar or different. The polarity-oriented approach was employed in this analysis, as my aim was to see whether the translation-equivalent pairs in question would occur in matching positive or negative contexts. This approach has also in fact been extensively used by several researchers who have conducted earlier contrastive studies of semantic prosody (see section 2.7).

48 translation-equivalent pairs were investigated. The results reveal that of the 48 pairs, 36 display the same semantic prosody. Of these 36 pairs, 5 share a positive semantic prosody, and 1 shares a negative semantic prosody. The remaining 30 pairs do not tend to co-occur with either positive or negative words, and are thus deemed to share a neutral semantic prosody.

Even when the pairs share the same positive or negative semantic prosody, they may differ in terms of semantic preference. The pair /phûukhon/ ‘people’, for example, share a negative semantic prosody, but have different semantic preferences. While /phûukhon/ has a semantic preference for words expressing the qualities of being

crowded and chaotic, ‘people’ has a semantic preference for illness. The matching pairs may also differ in terms of prosodic strength, or the degree of preference (see section 6.6), such that one word in the matching pair has a stronger prosodic strength than the other. For example, /mâattràthǎan/ and ‘standard’ both show a positive semantic prosody, but /mâattràthǎan/ has a higher positive prosodic strength than ‘standard’ (quantified as 8 vs. 4.7; see section 6.5.3).

There are 12 translation-equivalent pairs where the words in the pairs have a different semantic prosody. That said, there are no cases where one word in the pair has a positive semantic prosody, and the other has a negative semantic prosody. In every case, one word in the pair has a positive or negative semantic prosody, and the other has a neutral semantic prosody. This arguably suggests that there is a relationship between semantic prosody in Thai and English; otherwise we would expect at least some cases where one word in the pair has a positive semantic prosody, and the other has a negative semantic prosody. This relationship must arise from a functional basis, that is the meanings the words express, since the languages are genetically and areally unrelated, as noted in section 1.1.

Thus, in sum, the answer to my last research question is that while some pairs have the same semantic prosody, others do not. In cases where the pairs share the same semantic prosody, they may differ in terms of semantic preference and prosodic strength. Where differences are found, the pairs are arguably not “direct” equivalents, although they are still translation-equivalents (see discussion in section 6.6).

Some methodological issues arose from this analysis. We have seen that it is not always possible to classify collocates when they are considered out of context. This is

partly due to tokenisation errors, and also the fact that sometimes the evaluative meaning of the collocate depends on which sense it is being used in, what node it co-occurs with, and so on. It is thus sometimes necessary to look at the concordance lines where the collocates in question occur so as to classify those collocates accurately. This dependence on concordance lines underlines a problem of the polarity-oriented approach operationalised via statistical collocates, which is that it cannot always be used alone, but has to be complemented with at least some concordance analysis.

7.3 Limitations of the study

This study has some limitations in terms of the methods employed. The first limitation is concerned with tokenisation. The TNC is under development, so the tokenisation in the corpus is not always consistent. There are also actual tokenisation errors. Tokenisation errors seem to have particularly affected the polarity-oriented approach, as errors in tokenisation result in erroneously-tokenised collocates being identified statistically. These inaccurate collocates do not seem to have a major impact on the result of the analysis, as they are mostly erroneously-split parts of proper nouns, which are neutral in polarity. Erroneously-split parts of loanwords or simplex words were classified according to the meaning of the original word from which they are split (see section 6.5.2).

Nevertheless, there are also cases where the debatable and/or inconsistent tokenisation of complex words employed in the corpus does appear to affect the result of the analysis. We have seen this issue with the node /phǒnɲaan/ ‘achievement’ (see section 6.5.2), where I counted /boo/, which is neutral, rather than /boo dɛɲ/, which is positive,

as the collocate of the node. However, whether I counted /boo/ or /boo dəɛŋ/, the end result is still that /phǒnŋaan/ has a positive semantic prosody. Counting /boo/, rather than /boo dəɛŋ/, only affects the total numbers of positive and neutral collocates. There are three other cases like this one. In all these three cases, regardless of whether we count the collocate as it appears on the collocate list or the complex word from which that collocate is separated, the overall semantic prosody of the node is the same.

That said, in the absence of these tokenisation errors, the frequencies of each collocate type would be more precise. The tokenisation errors do also in fact affect the EUM-oriented approach, a point I will discuss later.

The second limitation is associated with my method for classifying collocates as positive, negative, or neutral. Throughout the analysis of the Thai data, I used my introspection as a native speaker of Thai to classify collocates as positive, negative, or neutral (see section 3.4.1.3). I conducted an inter-rater agreement study to justify the use of my introspection. The results of that study show that my introspection is mostly in agreement with that of other Thai speakers who acted as participants.

Nevertheless, it might be objected that the inter-rater agreement study I conducted is rather small relative to the much bigger actual analysis. In consequence, I might have classified some collocates inaccurately in my actual analysis, and this could in turn sway the semantic prosody of the node, as within the polarity-oriented approach, the criteria for semantic prosody are justified based on the proportion of positive, negative, and neutral collocates. This criticism is not invalid. However, I would argue that the identification of a semantic prosody is never based on one or just a few collocates. Rather, it is based on many collocates, and on certain specified criteria (see section 3.4.1.1). Therefore, it is

unlikely that, if I did happen to classify one or two collocates inaccurately, that would result in the wrong semantic prosody being reported for that node. In fact, for each node, there were only a few cases where I was not very sure of the correct classification.

A parallel limitation is associated with the classification of English collocates. Since I am not a native speaker of English, there were cases where I was not sure of the correct classification and needed to consult the dictionary or to seek an opinion from a native speaker of English. Thus, there might well be cases where I or the native speaker that I consulted classified the collocates inaccurately. However, as with the Thai collocates, it should be recalled that the identification of semantic prosody is based on a large number of collocates rather than on one or just a few collocates. Therefore, the inaccurate classification of one or two collocates ought not to sway the semantic prosody of the node under study.

Moreover, it is worth noting that to my knowledge, *all* researchers in the literature who examine semantic prosody, whether in English or other languages, and adopt the polarity-oriented approach, use their introspection as a native speaker to classify collocates. They do not conduct even limited inter-rater agreement studies. Therefore, it is safe to say that, to the extent that the lack of across-the-board inter-rater testing is a limitation of my study, it is a limitation that afflicts the entire field. My study is thus arguably methodologically somewhat more valid and, in this way at least, represents an improvement on previous studies.

The third limitation is concerned with the actual number of concordance lines analysed for each verb in Chapter 5. Even though I aimed to explore 200 concordance lines for each verb in each genre, there were not always 200 concordance lines available

in the data from each genre. This is mostly due to the tokenisation issues in the corpus. A verb within a nominalisation is tokenised separately from the other elements of the nominalisation. It thus appears as a separate unit, as if it was used as a verb, not part of the nominalisation. These instances of nominalisation were all excluded from my concordance analysis (see section 5.4), reducing the number of available concordance lines. There are also cases where concordance lines had to be excluded because of tokenisation errors. /mêε hōŋ sǎn/, for example, is the name of a province. This proper noun is tokenised in the current TNC as three units: /mêε/, /hōŋ/, and /sǎn/. Therefore, concordances for the verb /sǎn/ also include instances where /sǎn/ is actually part of /mêε hōŋ sǎn/, and not a verb. These instances were excluded from the analysis. Finally, there are repeated concordance lines. This is because there are duplicated texts in the TNC. Repeated concordance lines were only included once.

Due to the above reasons, there were not always 200 concordance lines available for the analysis. The disadvantage of this is that I might have missed some general patterns or extended units of meaning. Furthermore, if I had had 200 concordance lines to analyse, I probably would have had more examples for each extended unit and thus would have been able to be more confident in my analysis. For example, I earlier argued (see section 5.5.19) that /còp chiiwít/ seems to me as a native speaker of Thai to be an extended unit of meaning that has a clear pragmatic function of expressing premature or unexpected death. But this pragmatic function seems to operate only in some of the examples of /còp chiiwít/. Due to low number of available examples, I cannot state with confidence that /còp chiiwít/ is an extended unit of meaning in the Sinclairian sense. If I

had had more examples, I would probably have been able to argue for this extended unit of meaning with more confidence.

That said, it is in fact a common limitation in any corpus research that we cannot analyse everything in a corpus. It may be true that analysing more data tends to enable us to find more patterns, especially infrequent ones, but it is not practically possible to do an infinite analysis. Rather, we have to stop somewhere. This also applies to my study. Moreover, the fact that I might have missed some general patterns or extended units of meaning does not mean that the patterns or the extended units of meaning I did discover are inaccurate. It just means that I would likely have had more evidence for infrequent patterns if I had had more examples to look at.

It might be suggested that to avoid the above problem, I could have analysed more concordance lines for each verb in each genre, but looked at fewer verbs. However, I would not adopt this approach over my present one. Examining more concordance lines *might* provide more evidence for each extended unit of meaning as earlier discussed, but it would provide less *coverage*. That is, since I only had a finite amount of time, more concordance lines per verb would have to mean fewer verbs. Different verbs tend to participate in different patterns or extended units of meaning. Moreover, most researchers studying semantic prosody tend to look at just a few specific words. Sinclair (1998: 20; 2004: 30-35), for instance, pays attention only to just a few words, namely *naked eye* and *budge* (see section 2.3.2.1 and section 2.3.2.2). Similarly, Hunston (2007: 251-253) just looks at a few words, such as *cause* (see section 2.3.4.2). To my knowledge, no prior study has looked at as many words as I have done, using the concordance-based method.

This is a point in favour of my having adopted the current approach of analysing a large number of words, rather than looking at only a few verbs.

7.4 Contributions

Despite the limitations discussed above, I would argue that the study has made some definite contributions to the field. First, to my knowledge, this study is one of the very first studies of the Thai language to adopt the corpus-based approach (see section 1.2). In fact, it is arguably the first study able to make the most of the Thai National Corpus. The TNC as used in the study is the version that is implemented in CQPweb, which has search tools and functions unavailable in the earlier publicly accessible interface to the TNC. This is in itself a novel contribution.

Second, this study examines semantic prosody via *both* the polarity-oriented approach *and* the EUM-oriented approach. Most studies in the literature investigate semantic prosody using only one of the two approaches, depending on which scholar's model the study is following. The advantage of applying both approaches with the same data within a single study is that it allows us to arrive at a comprehensive understanding of the differences between the two approaches (see section 4.5). This in turn enables us to understand what each approach is useful for. To be specific, the polarity-oriented approach is useful for the investigation of a word's tendency to appear in a positive, negative, or neutral environment. The EUM-oriented approach, by contrast, is useful for the study of an extended unit of meaning and its pragmatic function in the Sinclairian sense. Applying the polarity-oriented approach to a large number of nodes has also made us aware of one limitation of this approach, in that it is not always possible classify

collocates considered in isolation. Rather, we need to look at concordance lines, as the evaluative meaning of the collocate is sometimes contingent upon the sense it is being used in, the node it co-occurs with, and so on (see section 6.5.2). All this is a specifically methodological contribution.

Third, this is the first study that investigates semantic prosody in the Thai language. Most studies of semantic prosody examine semantic prosody in English. There have been numerous studies that look at other languages, mostly European, such as Spanish, Italian, Portuguese, and Chinese. These studies often feature the comparison of semantic prosodies between English and the language in question (see section 2.7). Thus, the fact that my study also looks at semantic prosody in Thai on its own, with the contrastive study as only *part* of the study, is a major novelty of thesis (see section 1.1). In addition, the fact that the two approaches to semantic prosody can be applied without major difficulty to the Thai language has substantiated the cross-linguistic adequacy of existing linguistic theories in the area of semantic prosody. This in turn helps to expand the boundary of linguistic knowledge in the area of semantic prosody. The knowledge gained from this study will lead the way to fruitful later studies on stylistics, (critical) discourse analysis, sociolinguistics, language teaching, linguistic theory, and translation – each of which has been informed by the study of semantic prosody in English – in the Thai language. Pushing the field forwards in this way is a major contribution of my thesis. Particularly, the study of semantic prosodies of Thai-English translation-equivalents gives valuable insights to both translators and language learners. We have seen that whereas some pairs of translation-equivalents under study share the same semantic prosody, others do not. Moreover, in cases where the pair shares the same

semantic prosody, their prosodic strength may be different. This understanding will help translators and language learners become more cautious when selecting a translation equivalent. This is a major practical contribution of my thesis.

7.5 Possible future research

As I have noted previously, the TNC is under development. So far only around 32 million words have been added to the corpus. Once the corpus is fully developed, it will consist of around 80 million words, and hopefully, at the same time, problems in its tokenisation will have been resolved. Without tokenisation errors, many of the methodological limitations discussed in section 7.3 would be much reduced. Moreover, using the complete corpus, it would be possible to explore semantic prosody across a wider variety of genres, not just *academic*, *fiction*, *newspaper*, and *non-academic*, as there will be more data available for the analysis in each genre (see section 5.3). A potential direction for further research would thus be to repeat this study using the complete TNC, or to look at different genres.

In my analysis of variation in semantic prosodies across genres, I excluded all instances where the verb under study is nominalised (see section 5.4). It would therefore be interesting to look at these instances, to see to what extent nominalised verbs participate in the same or different patterns, or extended units of meaning, as when functioning as a verb, and whether there will exist variation across genres. Thus, another avenue for future research would be to take full account of nominalisation of verbs.

In my contrastive study of translation-equivalents, I was only able to look at one English translation-equivalent, which I identified using a bidirectional dictionary, in

comparison to each Thai word. It would be possible to look at more English translation-equivalents, so as to see which one of the suggested equivalents has the highest degree of equivalence in terms of semantic prosody, semantic preference, and prosodic strength to the Thai word under study. Analysing more English translation-equivalents for each Thai word in this way would be another direction for research.

I identified the English translation-equivalents from a bidirectional dictionary and employed two comparable corpora as the basis for my analysis. Some researchers in the literature who conduct contrastive studies (see section 2.7) instead use a bidirectional (or parallel) corpus for their identification of translation-equivalent pairs. Others also use a bidirectional corpus in their actual analysis of semantic prosody. Thus, another challenging research project would be to use a Thai-English bidirectional corpus, if there is one available, to identify translation-equivalent pairs and to investigate semantic prosody. If no such corpus exists, of course, creating one would itself be a major research undertaking.

Finally, I noted earlier that within the polarity-oriented approach, I do not adopt Louw's partly-implicit notion that semantic prosody is a result of a semantic transfer that takes place over time (see section 2.5.3). Rather, I consider semantic prosody within a synchronic framework. Clearly then, a possible area for future research would be to conduct analysis of semantic prosody, either in Thai alone or in multiple languages in comparison, using a diachronic corpus, to see whether or not that conceptualisation of semantic prosody as a diachronic phenomenon is reflected in corpus data over time.

7.6 Final remarks

Undertaking the research presented in this thesis has been one of the hardest tasks I have achieved in my life. It has been a very long journey fraught with tears. It might be thought that working with Thai should be an easy task for me, because it is my first language. I would say that it was in fact hard and challenging. Undoubtedly, as a native speaker, I can speak Thai fluently. However, when it comes to Thai grammar in particular, I would say that there are multiple areas which, when I started, I had no knowledge of. But now, after completing the thesis, I would say that as well as learning more about semantic prosody in Thai, which is the biggest achievement of this thesis, I have considerably learned more about the linguistic structure of Thai in general. Finally, I hope that this thesis, given its substantive contributions to the field as outlined above (see section 7.4), will pave the way for other researchers to take this area of study further.

References

- Aroonmanakun, W. 2007. 'Creating the Thai National Corpus', *MANUSYA: Journal of Humanities* 13: 4-17.
- Aroonmanakun, W., Tansiri, K. and Nittayanuparp, P. 2009. 'Thai National Corpus: a progress report', in *Proceedings of the 7th Workshop on Asian Language Resources*, pp. 153-60. Suntec, Singapore.
- Aston, G. and Burnard, L. 1998. *The BNC Handbook: Exploring the British National Corpus with SARA*. Edinburgh University Press.
- Aw, A., Aljunied, S.M., Lertcheva, N. and Kalunsima, S. 2014. 'TaLAPi - a Thai linguistically annotated corpus for language processing', in *Proceedings of Ninth International Conference on Language Resources and Evaluation*, pp. 125-32. Reykjavik, Iceland.
- Bahns, J. 1993. 'Lexical collocations: a contrastive view', *ELT Journal* 47 (1): 56-63.
- Baker, P. 2006. *Using Corpora in Discourse Analysis*. London: Continuum.
- Baker, P and Egbert, J. 2016. 'Introduction', in P. Baker and J. Egbert (eds.) *Triangulating Methodological Approaches in Corpus-Linguistic Research*. London: Routledge.
- Bednarek, M. 2008. 'Semantic preference and semantic prosody re-examined', *Corpus Linguistics and Linguistic Theory* 4 (2): 119-39.
- Berber-Sardinha, T. 2000. 'Semantic prosodies of English and Portuguese: a contrastive study', *Cuadernos de Filología Inglesa (University of Murcia, Spain)* 9 (1): 93-110.
- Biber, D. 1988. *Variation Across Speech and Writing*. Cambridge University Press.
- Biber, D. 1993. 'Representativeness in corpus design', *Literary and Linguistic Computing* 8 (4): 243-57.
- Biber, D., Conrad, S. and Reppen, R. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Biber, D., Conrad, S. and Cortes, V. 2004. 'If you look at...: lexical bundles in university teaching and textbooks', *Applied Linguistics* 25 (3): 371-405.
- Biber, D. and Conrad, S. 2009. *Register, Genre, and Style*. Cambridge University Press.
- Campbell, G. and King, G. 2013. *Compendium of World's Languages*. Oxon: Routledge.
- Carter, R. and McCarthy, M. 1988. *Vocabulary and Language Teaching*. Routledge.
- Chametzky, R. A. 2000. *Phrase Structure: From GB to Minimalism*. Blackwell.
- Chi, M. L., Wong, P. Y. and Wong, C. P. 1994. 'Collocational problems amongst ESL learners: a corpus-based study', in *Proceedings of Joint Seminar on Corpus Linguistics and Lexicology*, pp. 157-65. Language Centre, HKUST, Hong Kong.
- Dam-Jensen, H. and Zethsen, K. 2008. 'Translator awareness of semantic prosodies', *Target* 20 (2): 203-21.
- Department of Provincial Administration. 2015. Retrieved August 23, 2016, from http://stat.dopa.go.th/stat/statnew/upstat_age_disp.php
- Ebeling, S. O. 2013. 'Semantic prosody in a cross-linguistic perspective', *Studies in Variation, Contacts and Change in English* 13. Reprinted online <http://www.helsinki.fi/varieng/series/volumes/13/ebeling/>

- Ebeling, S. O. 2014. 'Cross-linguistic semantic prosody: the case of 'commit', 'signs of' and 'utterly' and their Norwegian correspondences', *Oslo Studies in Language* 6 (1): 161-79.
- Ellis, N. C. 2012. 'What can we count in language, and what counts in language acquisition, cognition, and use?', in S. T. Gries and D. Divjak (eds.) *Frequency Effects in Language Learning and Processing*, pp. 7-34. Berlin: Mouton de Gruyter. Reprinted online <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.472.2582&rep=rep1&type=pdf>
- Firth, J. R. 1957. *Papers in Linguistics*. Oxford University Press.
- Friginal, E. and Hardy, J. 2014. *Corpus-Based Sociolinguistics: A Guide for Students*. Oxon: Routledge.
- Garside, R. 1987. 'The CLAWS word-tagging system', in R. Garside, G. Leech and G. Sampson (eds.) *The Computational Analysis of English: A Corpus-Based Approach*, pp. 30-41. London: Longman.
- Hardie, A. (forthcoming) 'A dual sort-and-filler strategy for statistical analysis of collocation, keywords, and lockwords'.
- Hoey, M. 2004. 'Lexical priming and the properties of text', in A. Partington, J. Morley and L. Haarman (eds.) *Corpus and Discourse*. Bern: Peter Lang.
- Hoey, M. 2005. *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- Hoey, M. 2007. 'Grammatical creativity: a corpus perspective', in M. Hoey, M. Mahlberg, M. Stubbs and W. Teubert (eds.) *Text, Discourse, and Corpora: Theory and Analysis*. pp. 31-56. London: Continuum.
- Horrocks, G. C. 1987. *Generative Grammar*. London, New York: Routledge.
- Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge University Press.
- Hunston, S. 2004. 'Counting the uncountable: problems of identifying evaluation in a text and in a corpus', in A. Partington, J. Morley and L. Haarman (eds.) *Corpora and Discourse*, pp. 157-88. Bern: Peter Lang.
- Hunston, S. 2007. 'Semantic prosody revisited', *International Journal of Corpus Linguistics* 12 (2): 249-68.
- Hunston, S. and Francis, G. 2000. *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.
- Isahara, H., Ma, Q., Sornlertlumvanich, V. and Takahashi, N. 2000. 'ORCHID: building linguistic resources in Thai', *Literary and Linguistic Computing* 15 (4): 465-78.
- Iwasaki, S. and Ingkaphirom, P. 2005. *A Reference Grammar of Thai*. Cambridge University Press.
- Lakoff, G. and Johnson, M. 1980. *Metaphors We Live By*. University of Chicago Press.
- Lee, D. 2001. 'Genres, registers, text types, domains, and styles: clarifying the concepts and navigating a path through the BNC jungle', *Language Learning and Technology* 5 (3): 37-72.
- Leech, G., Garside, R. and Bryant, M. 1994. 'CLAWS4: the tagging of the British National Corpus', in *Proceedings of the 15th International Conference on Computational Linguistics (COLING 94)*, pp. 622-28. Kyoto, Japan.
- Löbner, S. 2002. *Understanding Semantics*. Oxon: Routledge.

- Louw, W. E. 1993. 'Irony in the text or insincerity in the writer? The diagnostic Potential of semantic prosodies', in M. Baker, G. Francis and E. Tognini-Bonelli (eds.) *Text and Technology: In Honour of John Sinclair*, pp. 157-76. Amsterdam: John Benjamins.
- Louw, W.E. 2000. 'Contextual prosodic theory: bringing semantic prosodies to life', in C. Heffer and H. Sauntson (eds.) *Words in Context: A Tribute to John Sinclair on his Retirement*, pp. 48-94. University of Birmingham.
- Mautner, G. 2007. 'Mining large corpora for social information: the case of elderly', *Language in Society* 36 (1): 51-72.
- McEnery, T., Xiao, R. Z. and Tono, Y. 2006. *Corpus-Based Language Studied: An Advanced Resource Book*. London: Routledge.
- McEnery, T. and Hardie, A. 2012. *Corpus Linguistics*. Cambridge University Press.
- Morley, J. and Partington, A. 2009. 'A few frequently asked questions about semantic - or evaluative - prosody', *International Journal of Corpus Linguistics* 14 (2): 139-58.
- Munday, J. 2013. 'Looming large: a cross-linguistic analysis of semantic prosodies in comparable reference corpora', in A. Kruger, K. Wallmach and J. Munday (eds.) *Corpus-Based Translation Studies: Research and Applications*, pp. 169-86. London: Bloomsbury Academic.
- Nesselhauf, N. 2005. *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.
- Partington, A. 1998. *Patterns and Meaning: Using Corpora for English Language Research and Teaching*. Amsterdam/Philadelphia: John Benjamin.
- Partington, A. 2004. "'Utterly content in each other's company": semantic prosody and semantic preference', *International Journal of Corpus Linguistics* 9 (1): 131-56.
- Partington, A. 2014. 'Evaluative prosody', in K. Aijmer and C. Rühlemann (eds.) *The Handbook of Corpus Pragmatics*, pp. 532-80. Cambridge University Press.
- Partington, A. (forthcoming) 'Evaluative clash, evaluative cohesion and how we actually read evaluation in texts'.
- Philip, G. 2011. *Colouring Meaning: Collocation and Connotation in Figurative Language*. Amsterdam: John Benjamins.
- Saeed, J. I. 1997. *Semantics*. Willey-Blackwell.
- Schmitt, N and Carter, R. 2004. 'Formulaic sequences in action: an introduction' in N. Schmitt (ed.) *Formulaic Sequences*, pp. 1-22. Amsterdam: John Benjamins.
- Sinclair, J. 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing*. London: Collins.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford University Press.
- Sinclair, J. 1998. 'The lexical item', in E. Weigand (ed.) *Contrastive Lexical Semantics*, pp. 1-24. John Benjamins.
- Sinclair, J. 2003. *Reading Concordances*. London: Longman.
- Sinclair, J. 2004. *Trust the Text: Language, Corpus and Discourse*. London: Routledge.
- Sinclair, J., Jones, S., Daley, R. and Krishnamurthy, R. 2004. *English Collocational Studies: The Osti Report*. London: Continuum.
- Smyth, D. 2002. *Thai: An Essential Grammar*. London: Routledge.
- Sornlertlumvanich, V., Takahashi, N. and Isahara, H. 1999. 'Building a Thai part-of-speech tagged corpus', *Journal of the Acoustical Society of Japan. (E)* 20 (3): 189-98.

- Stewart, D. 2010. *Semantic Prosody: A Critical Evaluation*. London: Routledge.
- Stubbs, M. 1995. 'Collocations and semantic profiles: on the cause of the trouble with quantitative studies', *Functions of Language* 2 (1): 23-55. Reprinted online <https://www.uni-trier.de/fileadmin/fb2/ANG/Linguistik/Stubbs/stubbs-1995-cause-trouble.pdf>
- Stubbs, M. 1996. *Text and Corpus Analysis: Computer Assisted Studies of Language and Culture*. Oxford: Blackwell.
- Stubbs, M. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Stubbs, M. 2009. 'The search for units of meaning: Sinclair on empirical semantics', *Applied Linguistics* 30 (1): 115-37.
- Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Tognini-Bonelli, E. 2002. 'Functionally complete units of meaning across English and Italian: towards a corpus-driven approach', in B. Altenburg and S. Granger (eds.) *Lexis in Contrast: Corpus-Based Approaches*, pp. 73-95. John Benjamins.
- Wei, N. and Li, X. 2013. 'Exploring semantic preference and semantic prosody across English and Chinese: their roles for cross-linguistic equivalence', *Corpus Linguistics and Linguistic Theory* 10 (1): 103-38.
- Whitsitt, S. 2005: 'A critique of the concept of semantic prosody', *International Journal of Corpus Linguistics* 10 (3): 283-305.
- Xiao, R. Z. and McEnery, T. 2006. 'Near synonym, collocation and semantic prosody: a cross-linguistic perspective', *Applied Linguistics* 27 (1): 103-29.
- Zhang, W. 2009. 'Semantic prosody and ESL/EFL vocabulary pedagogy', *TESL Canada Journal*, 26 (2): 1-12.
- Zhang, C. 2010. 'An overview of corpus-based studies of semantic prosody', *Asian Social Science*, 6 (6): 190-94.

Dictionaries

- Macmillan Dictionary | Free English Dictionary And Thesaurus Online*. Retrieved February 12, 2016, from <http://www.macmillandictionary.com/>
- Rohrer, J. 2007. *Thai English German Dictionary*. Bangkok: Windows on the World Publishing Company Limited.
- The Royal's Institute. 2011. *พจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 [The Royal's Institute Dictionary 2011]*. Bangkok: Nanmeebooks.
- Thiengburanatham, W. 2001. *SE-ED's Modern Thai-English Dictionary*. Bangkok: Se-Education Public Company Limited.

Appendix – 1 Phonology and writing System

A1.1 Phonology

The Thai sound system consists of 21 consonants including two glides, all of which can occur as an initial consonant, and which are listed in Table A1.1. Only the glides and seven of the other consonants can occur as a final consonant, however, as shown in Table A1.2.

	Labial	Alveolar	Palatal	Velar	Other
Stop					
voiced	/b/	/d/			
vl. unaspirated	/p/	/t/	/c/	/k/	/ʔ/
vl. aspirated	/ph/	/th/	/ch/	/kh/	
Fricatives					
voiceless	/f/	/s/			/h/
Nasal	/m/	/n/		/ŋ/	
Liquid		/l/			/r/
Glide	/w/		/y/		

Table A1.1 Initial consonants (after Iwasaki and Ingkaphirom 2005: 4)

	Labial	Alveolar	Palatal	Velar	Other
Stop	/-p/	/-t/		/-k/	/ʔ/
Nasal	/-m/	/-n/		/-ŋ/	
Glide	/-w/		/-y/		

Table A1.2 Final consonants (after Iwasaki and Ingkaphirom 2005: 4)

The occurrence of the glottal stop /ʔ/ can generally be predicted. It is always present when there is no initial consonant, as in /ʔaacaan/ ‘teacher’, and when there is no final consonant after a short vowel, as in /tóʔ/ ‘table’. Thus, in these circumstances, Iwasaki and Ingkaphirom recommend not transcribing the glottal stop. However, for consistency, I have always transcribed it in these circumstances throughout the analysis.

The presence of the glottal stop cannot be predicted after diphthongs, and in these cases, it is always marked.

There also exist initial consonant clusters. They are formed by certain voiceless stops with a liquid, /l/ or /r/, or the bilabial glide /w/. The 11 possible initial consonant clusters are shown in Table A1.3

/pr-/	/tr-/	/kr-/	/kw-/
/pl-/		/kl-/	
/phr-/		/khr-/	/khw-/
/phl-/		/khl-/	

Table A1.3 Initial consonant clusters (from Iwasaki and Ingkaphirom 2005: 5)

There are nine short vowels, all of which have a contrasting long vowel. The short vowels are transcribed with one letter, whereas the long ones are transcribed with two letters, as shown in Table A1.4

	Front	Central	Back
High	/i/ /ii/	/ɯ/ /uɯ/	/u/ /uu/
Mid	/e/ /ee/	/ə/ /əə/	/o/ /oo/
Low	/ɛ/ /ɛɛ/	/a/ /aa/	/ɔ/ /ɔɔ/

Table A1.4 Vowels (after Iwasaki and Ingkaphirom 2005: 5)

There are three pairs of diphthongs; each has short and long versions. The short version is followed by /ʔ/, as shown in Table A1.5.

/iaʔ/	/uaʔ/	/uaʔ/
/ia/	/ua/	/ua/

Table A1.5 Diphthongs (from Iwasaki and Ingkaphirom 2005: 5)

One important characteristic of Thai is that it is a tone language, in which the pitch at which a syllable is pronounced is phonologically relevant (Smyth 2002: 5). There are five different tones in Thai: mid tone (no accent), low tone (ˊ), falling (ˋ), high tone (ˊˊ), and rising tone (ˊˋ), as shown in Table A1.6.

1 st tone: Mid	(no mark)	a	/khaa/ ‘to be settled’
2 nd tone: Low	`	à	/khàa/ ‘galangal root’
3 rd tone: Falling	^	â	/khâa/ ‘to kill’
4 th tone: High	´	á	/kháa/ ‘to trade’
5 th tone: Rising	ˇ	ǎ	/khǎa / ‘leg’

Table A1.6 Five different tones in Thai (after Iwasaki and Ingkaphirom 2005: 5)

Thai also distinguishes between live syllables (/kham pen/) and dead syllables (/kham taay/). Dead syllables end with a stop consonant, /p, t, k, ʔ/. Live syllables are those that end with anything other than a stop consonant. The rules by which the writing system indicates tone are different for the two types of syllable.

A1.2 Writing system

Thai has a unique script, which originally evolved from one of the alphabets of South India (Smyth 2002: 11). The Thai script is close to the Lao and Cambodian scripts. The script is widely believed to have been devised by King Ramkhamhaeng (1780-1841), though this is controversial. The writing system is alphabetic. It is written across the page from left to right. There are no spaces between words; a space, when used, functions as a punctuation mark, like a comma or a full stop.

A1.2.1 Consonant letters

Thai has 44 consonant letters. They are arranged according to the traditional Indian alphabetic order, that is velar stops, palatals, dentals, bilabials, and finally sonorants (Smyth 2002: 11). The reason why the number of letters is higher than the actual number of consonants is that many consonants are represented by more than one letter. Originally, these letters represented additional consonants which are distinct in the

languages of India but which are not distinct in Thai. For example, the sound /th/ can be represented by six letters (ก ฏ ฏ ฏ ฏ ฏ), and the sound /s/ by four letters (ซ ศ ส ษ). Due to the small number of final consonant sounds, some consonant letters change their pronunciation when they occur as a final consonant. For instance, ก ฏ ฏ represent /th/ when they appear as an initial consonant, but /t/ when they occur in the final position. The Thai consonant chart is shown in Table A1.7.

ก	ข	(ข)	ค	(ค)	ฃ	ง	จ
ฉ	ช	ช	ฅ	ญ	ฎ	ฏ	ฐ
ฑ	ฒ	ณ	ด	ต	ถ	ท	ธ
น	บ	ป	ผ	ฝ	พ	ฟ	ภ
ม	ย	ร	ล	ว	ศ	ษ	ส
ห	ฬ	อ	ฮ				

Table A1.7 Thai consonant letters chart

The consonant letters are classified into three classes: *mid* (letters for unaspirated stops), *high* (some letters for aspirated stops and fricatives), and *low* (other letters for aspirated stops and fricatives, plus letters for nasals, liquids and glides).

This consonant letter classification, along with the type of syllable (live/dead) and the vowel, is important for the specification of tone by the writing system, according to the complex writing system, but is, however, not directly relevant in the present context.

A1.2.2 Vowel letters

Vowels are represented by diacritics. They can appear before, after, below, or above a consonant letter (Smyth 2002: 14). Table A1.8 lists the vowel letters and the sound they represent. The _ symbol indicates the position of the consonant.

Short vowels/diphthongs	Letters	Long vowels/diphthongs and others	Letters
/a/	๕ / ๖	/aa/	๕
/i/	๗	/ii/	๗
/u/	๘	/uu/	๘
/e/	๙	/ee/	๙
/ə/	๑๐	/əə/	๑๐
/ɛ/	๑๑	/εε/	๑๑
/o/	๑๒	/oo/	๑๒
/ɔ/	๑๓	/ɔɔ/	๑๓
/uaʔ/	๑๔	/ua/	๑๔
/iaʔ/	๑๕	/ia/	๑๕
/uaʔ/	๑๖	/ua/	๑๖
		/ay/	๑๗ / ๑๘
		/aw/	๑๗
		/am/	๑๗

Table A1.8 Vowel letters and the sound they represent (after Iwasaki and Ingkaphirom 2005: 23)

Appendix – 2 The questionnaire used in the inter-rater agreement study

	Word	Positive	Neutral	Negative	Not sure
1.	เดือดร้อน				
2.	ผลลัพธ์				
3.	เคลื่อนไหว				
4.	คู่สัญญา				
5.	ลูกตาม				
6.	หวั่นไหว				
7.	สิ่งแวดล้อม				
8.	ความเคลื่อนไหว				
9.	สันติ				
10.	ผล				
11.	ประโยชน์				
12.	เสมอภาค				
13.	รายได้				
14.	รบกวน				
15.	สมดุล				
16.	ผูกพัน				
17.	กล้า				
18.	กลัว				
19.	ชี้หน้า				
20.	ผาดโผน				
21.	วางท่า				
22.	ประพฤติ				
23.	เฮฮา				
24.	ชกต้อย				
25.	ปิ่นปาย				
26.	โอ้อวด				
27.	นินัย				
28.	อดิเรก				
29.	เก็บตัว				
30.	หาเรื่อง				
31.	เจ้าชู้				
32.	ชอบใจ				
33.	สังสรรค์				
34.	เบรี่ยว				
35.	รู้สึก				
36.	ผู้ใหญ่				
37.	ได้เปรียบ				
38.	เพลิตเพลิน				
39.	ว่างงาน				
40.	เปลี่ยนแปลง				

Appendix – 3 List of abbreviations used in line 2 of Thai examples

1	first person
2	second person
3	third person
AZP	adverbializing particle
CM	challengeability marker
COP	copula
COMP	complementizer
DAT	dative marker
DEM	demonstrative
DIR	directional auxiliary verb
HON	honorific
LINK	linker
LP	linking particle
NEG	negation/negative
NMLZ	nominalizer
PASS	passive
PFX	prefix
PL	Plural
POT	potential
PP	pragmatic particle
Q	question particle/marker
REC	reciprocal

REFL	reflexive
REL	relative
SBR	subordinator
SG	singular
SLP	speech level particle