

A Corpus-based Study of Thai and English Quantity Word Equivalents: *Lăay*, *Several* and *Many*

RAKSANGOB WIJITSOPON

Department of English and Corpus Linguistics for Digital Humanities Research Unit

Faculty of Arts, Chulalongkorn University, Bangkok, Thailand

Author email: raksangob.w@chula.ac.th

Article information	Abstract
<p>Article history: Received: 27 Jan 2021 Accepted: 15 Jun 2021 Available online: 28 Jun 2021</p>	<p>The present study investigates the Thai quantifier 'lăay' (หลาย) and its two major English lexical equivalents: 'several' and 'many', using data from an English-Thai parallel corpus, the Thai and British National Corpora. An examination of the parallel corpus reveals that the quantifier 'lăay' has a broad semantic property as it can express meanings related to small, medium and large quantities or just the plurality of entities. This provides support to an observation that the word can pose problems for Thai EFL learners, translators and interpreters when expressing the concept of 'lăay' in English. Based on the parallel corpus, two English quantity words that denote different scales of quantity, 'many' and 'several', were found to be among the most common lexical equivalents of 'lăay'. Further comparative investigation was conducted on the Thai and British National Corpora. It is found that 'lăay' and its two common equivalents have overlapping and different co-occurrence patterns, illustrating their lexical equivalent status and distinct usage profiles at the same time. Differences between the two English equivalents were then focused on so that empirical evidence of usage patterns of the two most common English lexical equivalents of 'lăay' can be obtained. Findings from the study provide an insight into phraseological patterns and pragmatic-discourse functions associated with the English equivalents, which Thai speakers of English can make use of as a possible framework for their decision making when translating 'lăay' into English.</p>
<p>Keywords: Quantity words Thai-English translation Thai-English lexical semantics Parallel corpus Comparable corpus Corpus-based translation <i>Lăay</i>, <i>several</i>, <i>many</i></p>	

INTRODUCTION

The present study originated from my personal observation when teaching translating Thai into English that the Thai quantifier *lăay* can be translated into a number of English words, e.g., *many*, *several*, *a few*, *numerous*, *multiple* and *varied*. These words are also presented as equivalents of the Thai word *lăay* in English-Thai and Thai-English dictionaries (e.g., Iamworrathem, 1995; Karnden & Wannapok, 2009; Tiangburanatham, 2008, 1995). More importantly, some of these equivalents can be considered as contradictory in their denotation of scales of quantity. For instance, *many* is considered a 'mutal' quantifier denoting a large number (Biber et al., 1999; Quirk et al., 1985) whereas *several* is considered a 'paucal' (Quirk et al., 1985) or 'low-magnitude' quantifier (Moxey & Sanford, 2000) that denotes a small number. Biber et al. (1999)

put *several* in the same group as *some* since they are treated as expressing a moderate quantity. Huddleston and Pullam (2002) compare *several* with a *few*, while Dichtel (2016) compares *many* with *a lot of* as each pair are considered to be more or less synonymous. These grammatical and semantic accounts suggest that expressing the concept of *lǎay* in English can be particularly challenging to Thai EFL learners, teachers, translators and interpreters; there are various equivalents of *lǎay*, as recorded in dictionaries, and choosing a particular word instead of other possible words, say *many* instead of *several*, can give a different representation of the state of affairs under discussion, as suggested by the semantic differentiation among English quantity words argued in the above references.

The present study therefore seeks to systematically examine lexical equivalents of *lǎay* that Thai speakers of English can resort to when dealing with the Thai quantifier, using a corpus linguistic approach. The following research questions are asked:

1. What are common English equivalents of *lǎay*?
2. To what extent and in what ways are *lǎay* and its common English equivalents similar and different phraseologically?

The first question addresses the quantitative aspect of the study. Although a number of English equivalents of *lǎay* have been given, as demonstrated above, it seems that there has been no reliable information about common equivalents of the word, which can suggest its tendency towards particular meaning associations in English. The second research question focuses on the qualitative aspect of linguistic description, specifically on similar and different ways in which common English equivalents of *lǎay* are used. Knowledge of these can enhance naturalness in translation and help Thai speakers of English opt for an appropriate equivalent of *lǎay* in their communicative context. These two research questions reflect a corpus linguistic approach to lexical items, in which frequency and recurrent patterns of words used in authentic texts are the main interest and qualitative descriptions of lexical items are formulated on the basis of their usage as evidenced by their observable frequencies and co-occurrence patterns, all of which are inseparable.

The article is organized as follows. An overview of quantity expressions is first given, followed by a methodology section which describes the data utilized and analytical approaches taken to the words under study. Subsequently, results are reported and discussed before concluding remarks are given.

QUANTITY EXPRESSIONS

Quantity expressions, especially quantifiers, have been studied extensively and from various perspectives (see, e.g., Badia, 2009; Carden, 1976; Hogg, 1977; Keenan & Paperno, 2012). From a grammatical perspective, quantifiers have often been described in terms of their grammatical properties, focusing on the countability of the nouns that are quantified (see, e.g., Quirk et al., 1985). In semantics, studies on quantity expressions look at varied phenomena, e.g., correspondence of quantified expressions to logical forms, ambiguity of propositional

representations of quantity and domain restrictions that assist in evaluation of the truth of the quantity representation (see, e.g., Jaszczolt, 2002). It has been argued, however, that there is a need to incorporate both semantic and pragmatic perspectives on expressions of quantity. This is because, for example, constituents of quantity representation can be “epistemically determined by the audience in the context” (Bach, 2000, p. 271 cited in Jaszczolt, 2002, p. 94). The focus on contextual factors is in line with Moxey and Sanford’s (2000) psycholinguistic experiment, which argues that even numerical expressions, which are both accurate and mathematically precise regarding an amount of something, can give rise to different meanings depending on the context. Moxey and Sanford (2000) illustrate this with the following examples: ‘contains 2% fat’, ‘contains only 2% fat’ and ‘98% fat free’. These different expressions of quantity are found to have different impacts on interpretation. If this is true for numerical expressions, it is surely even more the case for quantity words that express an inherently vague enumeration like *lăay*.

A pragmatic perspective on quantity expressions centres on the concept of vagueness. Communication of quantity in discourse is achieved mainly through numbers and non-numerical (or verbal) expressions, such as quantifiers and approximators. It has often been assumed that the verbal expressions of quantity are less precise, serving to make descriptive statements vague. This vagueness, however, has been argued to be “purposive vagueness” (Powell 1985) because the choice of quantifiers instead of numbers actually serves certain communicative functions (see, e.g., Altenberg, 1990; Channell, 1990; Drave, 2002; Swales, 1990). For example, paucal quantifiers are observed to be used for mitigation and to help avoid categorical statements. They are frequently classified as hedges or downtoners (Ruzaitė, 2009). Moreover, Moxey and Sanford’s (1993) psycholinguistic study shows that passages seem to be easier to recall when natural language expressions are used instead of numbers. Finally, Powell (1985) argues that the meaning of quantifiers usually encodes some evaluative content concerning the significance of a quantity, which would be lost if a precise expression was used instead. Quantifiers are thus a linguistic device for speakers to convey a judgement with respect to the significance of the quantity, or to confer salience upon a judgment of quantity. This position is shared by Dubois (1987), which looks at scientific presentation papers and slides, where precision is highly expected – and yet quantifiers are found to be quite common. Dubois argues that quantifiers perform a major rhetorical function in the presentation and popularisation of scientific experiments. Imprecision arising from the use of quantifiers helps to foreground other, more precise quantities which the experimenter considers important. This is observed when authors present background information to their studies, when they cite theories, they believe are erroneous, and when they make estimates from previous research – while making precise statements on their own studies by using numbers. Dubois therefore suggests that quantifiers should be added to a list of modality expressions, as they are used to express one’s views on the importance of states of affairs. Nevertheless, it has also been argued that not all quantifying expressions have an evaluative dimension; *a few*, *many*, *a large number* and *a great deal* are evaluative, whereas *some* is neutral (Powell, 1985). This discourse-pragmatic aspect of quantity words will also be explored in the present study from a corpus linguistic perspective, focusing on the pattern – meaning/ function relationship (see, e.g., Partington 1998, Tognini-Bonelli, 2001).

While quantity expressions have been approached from various perspectives, there have been few cross-linguistic studies of quantity expressions (Ruzaitė, 2009). Most studies on quantity expressions focus on those in particular languages, including Modern Hebrew, Italian, Japanese and Taiwan Mandarin (cf. Keenan & Paperno, 2012). Quantifiers in Thai have scarcely been studied. The existing work on Thai quantity words addresses the questions of their word classes, e.g., whether they are quantifiers or adjectives, their semantic classification and particularly their syntactic ordering (see, e.g., Deephuengton, 1992; Indrambarya, 2011; Prasithrathsint, 2010; Stein, 1981; Savetamalaya, 1989). For instance, it has been shown that, in Thai, different quantifiers can co-occur to quantify a noun or noun phrase, and that there seems to be a systematic order of quantifiers when different ones are used to qualify a single noun or noun phrase. The wrong order of quantifiers can result in unacceptability of a sentence (Indrambarya, 2011).

With respect to the semantics of Thai quantifiers, it has been discussed mainly in connection with the syntactic ordering of Thai quantifiers, i.e., the ordering of co-occurring quantifiers is argued to be tied to semantic types of the quantifiers. Indrambarya (2011, p. 56) proposes that Thai quantifiers can be semantically classified into five groups: (1) *one* quantifiers, which denotes the meaning concerning ‘one’ or ‘each’, e.g., *tèlǎʔ* ‘each’; (2) *incomplete* quantifiers, which express quantities that are more or less than expected, e.g., *ʔiik* ‘more’ and *khêe* ‘just’; (3) *estimated* quantifiers, which conveys estimated number, e.g., *raaw* ‘approximate’; (4) *near-complete* quantifiers, which suggest the speaker’s perception of the quantity toward the full amount, e.g., *tʔat* ‘almost’; (5) *excessive* quantifiers, which indicate that the referred quantities exceed the full amount, e.g., *kwàa* ‘over’. The word *lǎay* is mentioned only in passing as a member in the *estimated* quantifier group. Interestingly, unlike what is presented in bilingual dictionaries, Indrambarya (2011) assigns *a few* as an equivalent to *lǎay* while *several* to the reiteration of *lǎay*, i.e., *lǎaylǎay*, which can be used as a quantifier in Thai (Note that, as mentioned above, *several* and *a few* are regarded as comparable by Huddleston and Pullam (2002)). This suggests that, approached from an introspection-based perspective, *lǎay* is a highly vague quantity word with meaning complexity, since it can convey differing scales of quantity in Thai. This can pose difficulty for Thai speakers who want to express ideas involving quantity in English. The present study, therefore, seeks to investigate meanings and uses of *lǎay* empirically and its common English equivalents, using a corpus-based approach.

RESEARCH METHODOLOGY

The data used in the present study come from three corpora:

- (1) the English-Thai Parallel Concordance (henceforth ETPC) (Aroonmanakul, 2009), which can be accessed from <http://www.arts.chula.ac.th/~ling/ParaConc/>,
- (2) the British National Corpus (BNC), available at <https://cqpweb.lancs.ac.uk/bncxmlweb/>
- (3) the Thai National Corpus (TNC), accessible from <https://cqpweb.lancs.ac.uk/tncv1/>.

ETPC was used to answer the first research question, i.e., to discover common translation equivalents of *lǎay*. Data contained by the concordancer are a variety of English source texts

and their Thai translations. This means that retrieved entries of *lǎay* are translations of various English words, rather than the other way round. It cannot be denied that the given data does not directly reflect an interest in Thai-English translation taken in the present study. However, given that availability of comparative Thai-English data is limited, that the focus of present study is on lexical equivalents and that the concordancer contains a variety of texts and text types, with 377,053 parallel sentences, the ETPC makes it possible to retrieve English expressions regarded by professional translators as equivalent to *lǎay* in the context of translation. In other words, it can be used to explore what English words prompt the use of *lǎay* as their equivalent.

After common translation equivalents of *lǎay* were retrieved, some of the most frequent items were selected to be studied in terms of their phraseological patterns, meanings and discourse functions to explain why they might be appropriate for the translation of *lǎay*. Because the ETPC shows only parallel sentences in source texts and translated texts, it cannot provide thorough usage-based information for qualitative analysis and hence the BNC and TNC were consulted to answer research question 2. The BNC and TNC data were accessed via the CQPweb (Hardie 2012), a web-based corpus analysis system that combines ease of use, power in corpus analysis software, flexibility and compatibility with any corpus. The corpus analytical functions available on the CQPweb are, for instance, frequency lists, distribution tables, collocations, concordancing and keywords. It should be noted that the TNC, whose development is in progress, is designed to be comparable to the hundred-million-word BNC (Aroonmanakul, 2007). However, at the moment, it has reached the size of 32 million words, consisting of written texts; there is no transcribed spoken data, and speech is represented only secondarily within written texts, e.g., interviews presented in magazines. For the purpose of the present study, automatic collocation extraction and concordance line analysis were conducted to formulate qualitative descriptions of patterns, meanings and functions of the quantity words.

The qualitative analysis was conducted in two major steps. First, a comparative analysis of *lǎay* and its most common English lexical equivalents was carried out, focusing on their statistically significant collocates, which allow us to see patterns of word uses that in turn point to their overlapping and different meaning associations when used in communication. Then, differences between the two English equivalents were focused upon, with particular attention to their discourse-pragmatic functions as evident across concordance lines. Because different methodologies, i.e., frequency information, automatic collocation extraction and concordance lines, were adopted to answer the research questions, for the sake of clarity, how each of the descriptive tools was employed in the study will be described when results are reported and discussed in corresponding sub-sections.

RESULTS AND DISCUSSION

This section reports on and discusses findings from analysis of (1) frequency information of English equivalents of *lǎay* and (2) collocations and discourse-pragmatic functions of *lǎay* and its two common English lexical equivalents, respectively.

1. Common translation equivalents of *lǎay*

A total of 200 concordance lines that contain *lǎay* as a quantifier in ETPC were studied. Table 1 below presents common English expressions and linguistic categories that prompt the use of the quantifier *lǎay* in Thai translations.

Table 1
English equivalents of *lǎay* in ETPC

Equivalent of <i>lǎay</i>	Frequency	Per cent
<i>many</i>	70	35
Plural nouns with no quantifiers	53	26.5
<i>several</i>	32	16
Plural numerals	11	5.5
Adjectives (e.g., <i>numerous</i>)	9	4.5
Adverbs (e.g., <i>repeatedly</i>)	9	4.5
<i>a few/ the next few/ a couple + of</i>	6	3
<i>a lot/a series/a variety/ lots/ loads/ plenty + of</i>	6	3
<i>most</i>	2	1
<i>some</i>	2	1
TOTAL	200	100

Based on the table above, it can be seen that various English quantifiers and noun phrases that denote quantity lead to the use of *lǎay*. First, English quantity expressions that denote a large number or variety are frequently translated into *lǎay*, as can be seen by the highest frequency of *many*, such formulaic expressions as *a lot/lots/loads of*, *a series/variety of*, adjectives, such as *different*, *numerous* and *multiple*, and adverbs like *often*, *one and another*, and *repeatedly*. However, it is interesting that those that are generally seen as indicating a small and moderate quantity are also expressed as *lǎay* in Thai translations, e.g., *a couple*, *several* and *some*. Furthermore, English plural nouns and numerals that do not suggest a large or small quantity are also often translated into *lǎay*; the use of plural nouns without quantifiers is the second most common equivalent, joined by plural numerals, e.g., *tens*, *hundreds* and *thousands*, which comes fourth on the list. Examples of the different kinds of equivalents from the parallel corpus are given below.

- (a) I have been coding for a couple of years.

ฉันได้ฝึกเขียนโค้ดมาหลายปีแล้ว

chǎn dǎay fùk khǎan khóot maa lǎay pii léew

- (b) The introduction of Buddhism and its subsequent acceptance by the Japanese court resulted in the submergence of Shinto, the native religion, for many years.

การนำพระพุทธศาสนาเข้าไปสู่ประเทศญี่ปุ่นและการที่ราชสำนักญี่ปุ่นยอมรับนับถือพระพุทธศาสนา ได้มีผลทำให้ชินโตซึ่งเป็นศาสนาพื้นเมืองตกอับไปเป็นเวลาหลายปี

kaan nam phrá phút thá? sàat sa nǎa khâw pay sùu pra thêet yîi pùn lé? kaan thîi rât chá sǎm nák yîi pùn yowm ráp náp thǔw phrá phút thá? sàat sa nǎa dǎay mii phǎn tham hây chin too sǔn pen sàat sa nǎa phǔwun mwan tòk ?àp pay pen wee laa lǎay pii

(c) It's been years since I've seen you!

เป็นเวลาหลายปีแล้วตั้งแต่ที่ฉันได้เจอคุณ!

pen wee laa lăay pii léew tâŋ tɛɛ thii chăŋ dâay cəə khun !

All these suggest that the Thai quantity word *lăay* has a broad meaning association, capable of expressing both mere plurality and different scales of quantity, with a slight tendency towards a large quantity. This in turn points to challenges in translating it into English.

Nevertheless, on the basis of the above frequency information, it seems that the three most common equivalents in Table 1, whose frequencies are over 10% and altogether account for over 70% of occurrences of *lăay*, can be particularly useful options for Thai EFL learners, translators and interpreters. The problem, however, is that they denote different concepts of quantities. The less problematic case seems to be the use of plural nouns without any quantifiers as they serve to generalize entities or simply suggest the plurality of entities involved, and hence the question of how much/many is not a major concern. This can be illustrated by the examples below.

(d) There are worse things in the world than being a Scarecrow.

ในโลกนี้มีหลายสิ่งที่ย่ำกว่าการเป็นหุ่นไล่กา

nay lōok nii mii lăay siŋ thii yɛɛ kwàa kaan pen hùn lây kaa

(e) It used to take months to reach the Antarctic.

มันต้องใช้เวลามากกว่าจะถึงทวีปแอนตาร์กติกา

man tōŋ chây wee laa lăay dəon kwàa cà? thũŋ thá wīp ?ɛn tàak tìk

(f) Her jeans have holes in them.

กางเกงยีนส์ของเธอมีรูหลายรูบนนั้น

kaŋ keeŋ yiin khǎwŋ thəə mii ruu lăay ruu bon nán

On the other hand, the two common lexical equivalents, *many* and *several*, can pose problems among Thai speakers of English because, as noted earlier, *many* is in the mutal group of quantifiers while *several* in the paucal group. How can we know, then, when to use *many* or *several* when translating *lăay* into English? This is a question to be explored in the next section.

2. Patterns and functions of *lăay*, *many* and *several*

Following the quantitative information in 1, in this section, collocational patterns of *lăay*, *many* and *several* were first investigated via the TNC and BNC to further our understanding of their lexical equivalent status from a qualitative perspective. As will be shown, there are overlapping and different co-occurrence patterns among the three words. Because of this, differences between the two English equivalents are further pursued in terms of their discourse functions, as can be inferred from the observed patterns in concordance lines.

2.1 Collocational patterns of *lăay*, *many* and *several*

Automatic extraction of statistically significant collocates of the three words under study was employed and the collocates were ranked by Log Ratio values (Hardie, forthcoming). According

to the CQPweb, the Log Ratio statistic measures how big the difference is between the (relative) frequency of the collocate alongside the node, and its (relative) frequency in the rest of the corpus. In this way, the Log Ratio statistics is very similar to the Mutual Information measure, which is one of the commonly used collocational measures, in that both measure effect size. However, Log Ratio is combined with a statistical significance filter, i.e., Log Likelihood. This means that collocates extracted via the CQPweb are included in the list if they are significant at the 5% level ($p < 0.05$).

The thresholds set for the extraction were as follows: a minimum frequency of the collocates to occur is 20, with a Log Ratio of at least 3. If a word occurs in less than 10 texts, it was excluded from the analysis. The default collocational span used on the CQPweb was opted for, i.e., three words to the left and right of the search word. Table 2 below demonstrates collocates of the three quantity words extracted on the basis of these criteria.

Table 2
Statistically significant collocates of *lăay*, *many* and *several*

No.	Collocates of <i>lăay</i>	Log Ratio	Collocates of <i>many</i>	Log Ratio	Collocates of <i>several</i>	Log Ratio
1	ชั่วคน <i>chũa khon</i> 'generations'	7.266	respects	5.652	occasions	6.138
2	พันปี <i>phan pii</i> 'thousand years'	5.367	facets	5.506	centimetres	5.517
3	ครา <i>kra</i> 'times'	5.258	commentators	5.112	kilometres	5.366
4	หน <i>hõn</i> 'times'	5.159	instances	5.1	times	5.199
5	เท่าตัว <i>thâw tua</i> 'times – multiply'	4.977	unanswered	4.835	layers	5.119
6	คลบ <i>tà lõp</i> 'times'	4.664	admirers	4.604	decades	4.971
7	กินเวลา <i>kin wee laa</i> 'last; take time'	4.221	ways	4.458	generations	4.971
8	ประการ <i>pra kaan</i> 'items to consider/ discuss'	4.087	contemporaries	4.362	thousand	4.746
9	แขนง <i>khà nǣŋ</i> 'branches'	3.91	observers	4.241	hundred	4.641
10	กรณี <i>kra thoŋ</i> 'law cases'	3.905	cases	4.208	respects	4.607
11	หลากหลาย <i>lâak</i> 'varied; various'	3.851	similarities	4.066	seconds	4.599

12	ทั้ง <i>thǎŋ</i> 'all; both'	3.813	thousands	4.039	weeks	4.426
13	ร้อย <i>rǔwɔy</i> 'hundred'	3.781	sociologists	3.994	centuries	4.339
14	ชนิด <i>chá nít</i> 'kinds; types'	3.583	varieties	3.991	phases	4.278
15	แง่มุม <i>ŋɛɛ mʉm</i> 'aspects; respects'	3.58	Englishmen	3.966	yards	4.129
16	หลาย <i>lǎay</i> 'several; many'	3.507	feminists	3.837	dozen	4.115
17	ชั้น <i>khǔm</i> 'levels; layers'	3.506	attractions	3.768	attempts	4.1
18	สิบ <i>síp</i> 'ten'	3.464	occasions	3.766	incidents	4.068
19	มากมาย <i>mâak maay</i> 'many'	3.445	alienated	3.749	hours	4.036
20	กิโล <i>ki? loo</i> 'kilograms; kilometres'	3.388	economists	3.746	advantages	4.031
21	อาทิ <i>ʔaa thi?</i> 'for example'	3.335	localities	3.727	months	4.008
22	เผ่าพันธุ์ <i>phǎw phan</i> 'species; ethnic groups'	3.332	insects	3.625	kinds	3.98
23	ครั้ง <i>khraŋ</i> 'times'	3.235	Christians	3.611	ways	3.964
24	พันล้าน <i>phan láan</i> 'billion'	3.187	varied	3.606	restaurants	3.948
25	ชาติพันธุ์ <i>cháat ti? phan</i> 'ethnic groups'	3.174	battles	3.596	dollars	3.944
26	ยี่ห้อ <i>yí hǔw</i> 'brands'	3.152	countries	3.532	pairs	3.928
27	ชิ้น <i>chín</i> 'pieces'	3.096	aspects	3.513	lasted	3.922
28	เชื้อชาติ <i>chúaa cháat</i> 'races'	3.076	anglers	3.494	instances	3.913
29	มิติ <i>mi? ti?</i> 'dimensions'	3.073	times	3.493	factors	3.91
30	หมื่น <i>mùwun</i> 'ten thousand'	3.064	scholars	3.491	lasting	3.853

Based on the threshold above, 30 collocates of *lǎay* have been yielded. On the other hand, the number of collocates of *several* and *many* are far more than 30, both having 82 collocates in total. The unequal number of collocates of *lǎay* and its English equivalents under the same criteria is due to the size of the two corpora. BNC is much bigger, so more collocates pass the significant test and are included on the list. At the same time, it also indicates that the minimum setting at 20 excludes more words in the TNC than the BNC, due to the size of the corpora. For a comparative purpose, the analysis in this section focuses on the top 30 collocates of the three words in question, although some of those below the top 30 ranking in the collocational lists of *many* and *several* will also be taken into account where relevant, since they also pass the significance test.

While reference grammars and textbooks often discuss differences among quantity words in terms of their scales and the countability of the nouns they quantify, the collocational approach to these words reveals that uses of quantity expressions can be compared and contrasted in terms of collocates and semantic preference patterns as well. According to Table 3 below, some collocates are shared by all three quantity words, i.e., *occasions*, *times* and *respects*. On the one hand, these collocates show how much the three words can overlap, despite differences in languages and basic semantics. On the other hand, they are problematic for purposes of differentiating *many* from *several*, and both from *lǎay*.

Apart from similarities shared by individual collocates, a number of collocates of the three quantity words can be seen as having semantic connections. Table 3 below presents the collocates of each word in different semantic groups. Within each group, collocates are ordered by their Log Ratio scores.

Table 3
Semantic groups of collocates of *several*, *lǎay* and *many*

Semantic group	<i>several</i>	<i>lǎay</i>	<i>many</i>
Occasion when something happens	occasions, times, phases	ครา <i>kra</i> 'times' ทน <i>hǎn</i> 'times' ตลอด <i>tà lǎp</i> 'times' ครั้ง <i>khráŋ</i> 'times'	times, occasions
Number	thousand, hundred, dozen	ร้อย <i>rǔay</i> 'hundred' สิบล <i>sip</i> 'ten' พันล้าน <i>phan láan</i> 'billion' หมื่น <i>mùun</i> 'ten thousand'	thousands
Generalised referential expressions	respects, advantages, kinds, ways, factors	ชนิด <i>chá nít</i> 'kinds; types' แง่มุม <i>ŋêe mum</i> 'aspects; respects' ชั้น <i>khum</i> 'levels; layers' มิติ <i>mít tít</i> 'dimensions'	respects, facets, ways, cases, similarities, varieties, aspects

Concrete items and classifier nouns	Restaurants, dollars, pairs	<p>ประการ <i>pra kaan</i> 'items to discuss'</p> <p>แขนง <i>khà nĕĕŋ</i> 'branches'</p> <p>กระทรวง <i>kra thoŋ</i> 'law cases'</p> <p>ยี่ห้อ <i>yīi hĕw</i> 'brands'</p> <p>ชิ้น <i>chĭn</i> 'pieces'</p>	Attractions, insects, localities, countries
Duration	decades, generations, seconds, weeks, centuries, hours, months, lasted, lasting	<p>ชั่วคน <i>chĭa khon</i> 'generations'</p> <p>พันปี <i>phan pii</i> 'thousand years'</p> <p>กินเวลา <i>kin wee laa</i> 'last; take time'</p>	
Measurement	Centimetres, kilometres, yards, layers	กิโล <i>ki? loo</i> 'kilograms; kilometres'	
Actions and events	attempts, incidents, instances		instances, battles
Variety		หลากหลาย <i>lĕak</i> 'varied; various'	varied
Quantity expressions		<p>เท่าตัว <i>thĕw tua</i> 'times – multiply'</p> <p>ทั้ง <i>thĕŋ</i> 'all; both'</p> <p>หลาย <i>lĕay</i> 'several; many'</p> <p>มากมาย <i>mĕak maay</i> 'many'</p>	
People			commentators, admirers, observers, contemporaries, sociologists, anglers, feminists, scholars, Englishmen, economists, Christians
Ethnicity		<p>เผ่าพันธุ์ <i>phĕw phan</i> 'species; ethnic groups'</p> <p>ชาติพันธุ์ <i>chĕat tĭ? phan</i> 'ethnic groups'</p> <p>เชื้อชาติ <i>chĭwa chĕat</i> 'races'</p>	
Miscellaneous		<p>อาทิ <i>?aa thĭ?</i> 'for example'</p>	unanswered, alienated

Beyond individual collocates, the three words have in common four semantic groups of collocates: (1) Occasions when something happens, (2) Numbers, (3) Generalized referential expressions, and (4) Concrete entities and classifier nouns. But even here, a look at the collocates within each group sheds light on slight differences among the three words. A case in point is the 'Number' group, in which only the word *thousands* is among the top 30 collocates of *many*, although its full collocate list also includes *millions* (LR 3.494/No. 33), *hundreds* (LR 3.159/No. 59), and *tens* (LR 3.048/No. 70). On the other hand, the top 30 collocates of *several* and *lĕay*

include more and varied numbers in the top 30 list, ranging from *dozen* to *billion*. This difference distinguishes *many* from *several* and *lǎay*; the latter two words are used to quantify numerical expressions more prominently than *many*, probably because *many* explicitly denotes a large number. In this regard, the use of *lǎay* can be seen as relatively closer to that of *several*.

Apart from the semantic preferences shared by all three words, a number of notional groups of collocates are shared by *lǎay* and either *several* or *many*. First, *lǎay* and *several* share two groups of collocates: ‘Measurement’ and ‘Duration’; these groups do not appear among the top 30 collocates of *many*. With respect to the ‘Measurement’ group, *lǎay* and *several* are very closely similar. None of the significant collocates of *many*, even those below the top 30, denote this concept. We may thus posit that both *lǎay* and *several* have specific functions in quantifying measurement. Moreover, for both words, these ‘Measurement’ collocates occur in two similar major phraseological constructions: (1) *lǎay/several* + measurement word (e.g., kilometres) and (2) *lǎay/several* + number (e.g., dozen/hundred/thousand) + measurement word. For both *lǎay* and *several*, the former phrasing occurs more frequently than the latter.

With regards to the ‘Duration’ group, the similarity between *lǎay* and *several* is not as clear. While this is the largest grouping of the top 30 collocates of *several*, no words in this group are among the top 30 collocates of *many*, and only two appear on its full list of 82 collocates: *centuries* (3.364/No. 41) and *years* (3.262/No. 49). It can therefore be stated that ‘Duration’ is another distinctive semantic preference of *several*, when compared with *many*. However, while the time collocates of *several* include *weeks*, *hours* and *months*, only terms denoting a particularly long duration, i.e., *ชั่วคน chûa khon* (generations), *พันปี phan pii* (thousand years) are significant collocates of *lǎay*. Intuitively, *lǎay* can also be used with nouns referring to shorter times – but such combinations do not turn up as statistically significant in the TNC, based on the criteria for collocation used in this study. In this case, the collocational profile of *lǎay* related to its function to quantify time duration can be seen as similar to both *several* and *many*, though in different ways; *lǎay* seems close to *several* in that both collocate in particular with words related to time and duration but the statistically distinctive use of *lǎay* to quantify time, like *many*, is particularly oriented towards long time periods, rather than also frequently occurring alongside time units of more modest magnitude, like *several*.

With regards to *lǎay* and *many*, the ‘Variety’ group displays a similarity shared by both words. The collocation between *lǎay* and *lǎak* forms idiomatic expressions in Thai, *lǎak lǎay* and *lǎay lǎak*, both of which express meanings about ‘Variety’; this is equivalent to the collocate *varied* of *many*, as in: “Despite many and varied flurries of activity in the first 40 minutes, Arsenal seemed reasonably secure if not inspired” (A8N). This collocational finding corresponds to Dichtel (2016), which uses corpus data to examine differences between the synonyms *many* and *a lot of* and has found that the former is different from the latter because its use highlights the concept of variety in particular. In this regard, *lǎay* and *many* are similar in that they are part of the unit of meaning (cf. Sinclair 1999) expressing the idea that something exists both in a large quantity and variety. Actually, *several* also has a collocate related to variety, i.e., *different*, but it occurs at No. 78 out of 82 on the list, with the Log Ratio value of 3.051. In other words, this collocational pattern and associated function is not prominent for *several* as it is for *many* and *lǎay*.

The remaining collocational groups are idiosyncratic for each word under study. First, the largest group of collocates on the top 30 list of *many* deals with ‘People’ in terms of their profession, specialization or interests. Intuitively, we might expect such nouns to collocate with any quantifier, but they occur significantly with *many* and not with *several*. Even further down the list, only three people nouns are found for *several*: *authors* (3.623/No. 39), *witnesses* (3.543/No. 46) and *females* (3.446/No. 48). Interestingly, people nouns do not occur as a significant collocate of *lǎay* in the present study, even though, intuitively, it seems to be used very often to quantify people nouns in Thai, such as *lǎay khon*. One of the reasons for this is related to the choice of the statistical measure adopted here. With the Log Ratio, the common people noun in Thai, *khon*, which can be used as a general noun and noun classifier, do not pass the statistical-significance filter on the CQPweb. In other words, the difference between the relative frequency of *khon* alongside *lǎay* and that of *khon* in the rest of the corpus is not big enough to be significant as a distinctive collocate of *lǎay*. We can thus say that on the basis of statistical measurement of collocation strength and significance test, *many* has a distinctive semantic preference for people nouns while *several* and *lǎay* do not.

Lǎay also has its distinctive semantic preference patterns. The ‘Quantity’ collocates can be regarded as a dominant semantic preference that distinguishes *lǎay* from its two English equivalents. The collocational extraction here suggests that the Thai quantifier is significantly used in collocation with other Thai quantifiers, including with the word *lǎay* itself. This hints at its broad semantic associations as observed earlier. In context, however, it can be observed that co-occurrences between *lǎay* and the ‘Quantity’ collocates, constituting formulaic quantity expressions in Thai, are often used for emphatic or intensifying purposes, as illustrated in this example:

โดยใช้ชีวิตเข้าแลกหลายครั้งหลายหนอย่างกล้าหาญและอดทน

dooy cháay chii wít khâw lǎek lǎay khráŋ lǎay hǒn yàaŋ klâa hǎan lé? ʔòt thon

‘He used his life in exchange [for his country] several/ many times, bravely and patiently.’

Here, the word *lǎay* is repeated in the formulaic expression *lǎay khráŋ lǎay hǒn*, wherein *khráŋ* and *hǒn* are synonymous, meaning ‘times’. In turn, this points to a tendency of *lǎay* to be reiterated and used idiomatically for an intensifying purpose, a pragmatic potential close to *many*.

Apart from the collocates concerning ‘Quantity’, another collocational group that is characteristic of *lǎay* only is that which consists of three words related to the concept of ethnicity: *phàw phan* (ethnicity), *châat tì? phan* (species or ethnic groups) and *chúua châat* (races). It is not clear why these three words, which are very closely related, are top 30 collocates of *lǎay* in the TNC, as they are relatively topic-specific. An overall interpretation could be given, i.e., that Thai discourse on ethnicity tends to feature the concepts of plurality and variety to a significant level. The multi-meaning *lǎay*, therefore, collocates significantly with these words.

Finally, the absence of the ‘Actions and Events’ collocational group can be seen as another difference between *lǎay* and its two common English equivalents. This is probably because the nouns denoting actions and events can be quantified in English but not in Thai, in which

the word *khraŋ*, meaning ‘times’, is quantified through *lǎay* instead.

In sum, then, it seems that the collocational patterns of *lǎay* seem to overlap with those of both *several* and *many* in different ways. This corresponds to the quantitative findings gained from the ETPC in Section 1 that *many* and *several* are among the most common lexical equivalents of *lǎay*: there are some collocates and four semantic preferences that are shared by all three words; *lǎay* and *several* share semantic preferences for ‘Time duration’ and ‘Measurement’ in particular while *lǎay* shares the semantic preference related to ‘Variety’ with *many*. At the same time, the comparative cross-linguistic collocational analysis has shown co-occurrence patterns that might furnish usage evidence on the distinction between the two English equivalents of *lǎay*, which has generally been dealt with in terms of perception of scale. On the basis of statistical measurement of collocation strength and significance test here, *many* has a distinctive semantic preference for variety and people nouns while *several* for time duration and measurement.

That said, it would be risky to treat the differences in semantic preference as an absolute differentiation between the two English equivalents to identify usage similarities between either of them and *lǎay*. This is because time and people nouns are common general nouns in English that *can* be quantified by both these quantity words (and others too, of course). There is evidence, for example, of *many years* (30 tokens per million words), *several people* (2.3 tokens per million words) and *lǎay khon* (142.23 tokens per million words). Rather, we should interpret the ‘preferred’ patterns as a signal of these English quantity words’ tendency to be used for particular communicative purposes in discourse, a point that can benefit Thai speakers of English when expressing the concept of *lǎay* in communication. This is discussed in detail in Section 2.2 below.

2.2 Discourse-pragmatic functions of *many* and *several*

To explore an observation based on collocational analysis above and further an insight into uses of *many* and *several* as possible equivalents of *lǎay*, in this section, three phraseological patterns are examined via concordance analysis as case studies of discourse functions of the two English quantifiers: (a) *several* + time duration nouns (b) *many* + people nouns and (c) *several times* vs. *many times*. Patterns (a) and (b) were studied because they are idiosyncratic semantic preference patterns of *several* and *many*, respectively, and hence it will be useful to see why the two quantifiers are remarkably used in collocation with the different semantic domains even though they seem intuitively possible for quantifying both time and people nouns. As for pattern (c), because both quantity words are used significantly with the noun *times*, comparing their collocation with the same word will shed light on their differing discourse-pragmatic functions in communication.

(a) *several* + time duration nouns

Upon an examination of concordance lines for the collocation between *several* and significant collocates in the ‘Duration’ group, it is found that the phrase mainly serves to provide vague temporal background information for states of affairs. As noted by text-linguistic and discourse

analysis scholars (cf. e.g., Gavins, 2007), time expressions are essential text-world building elements that provide deictic cues for readers/speakers to connect themselves with the text and discourse at hand. The quantifier *several* therefore constructs a moderate timeline of actions/events under discussion, as illustrated below.

The releases mark the first improvement in the human rights situation in Malawi for several years. (A0361)

We spent several days in the Egyptian museum. (FEM427)

(b) *many* + people nouns

A look at the concordance lines of the collocations between *many* and significant people noun collocates (see Table 3) reveals that the collocational pattern tends to be followed by cognition and speech verb phrases, as illustrated by Figure 1 below. This helps enhance the significance of the presented views and reactions, expressing that they are not simply the opinions of one or a few experts but rather are supported by a large number of them. *Many* therefore serves to construct an authoritative voice in the discourse since a vague reference to a large number of people who support some asserted view can raise the perceived validity or reliability of the asserted views, without the speaker or writer having to commit to the truthfulness of any exact number.

result in a reduction in capacity, leading	many observers	to criticise BR for short-sightedness in
exchange rate may pose a dilemma as	many economists	believe the economy is now poised to
ands while retaining its basic strength .	Many anglers	are hailing it as the most significant
index hit a post crash high of 2,423.9 .	Many observers	believe it will fall to below 2,200 points
than a year , were also broken off . But	many commentators	believe that there are points of
, again threatened the BBC 's position .	Many observers	were disturbed by the way in which the
not Harriet , was the heartbreaker . '	Many scholars	have suggested that she broke his heart
d a housewife 's weekly budget (which	many economists	dismiss as pre-Keynesian , naïve , and
rnment. This is significant , given that	many commentators	in 1983 thought that the policies of the
tion to the women 's movement. What	many feminists	find most surprising about the article is
orm. If it does not follow that fashion ,	many observers	fear the country will not sustain its usua
is labour-market reform . Back in 1983	many economists	were sceptical about the wage accord
ellent signs of industrial restructuring .	Many economists	say bankruptcy figures in Poland are
ming the fastest-growing in the world ,	many commentators	predicted that these great families
efore change if the situation changes .	Many sociologists	have been reluctant to employ the noti

Figure 1 Sample concordance lines of the phraseology of *many* + 'specialist' nouns

The people noun collocates with embedded meanings related to time and countries like *contemporaries* and *Englishmen* (see Table 3) are found to co-occur noticeably, though not always immediately adjacent to, with lexical signals of comparison and contrast (12 out of 31, or 38.71 per cent), e.g., *as*, *unlike*, *more*, *while* and *though*, suggesting that this group of people noun collocates is involved in comparative discourse, usually between different groups of people in the context of time or place. The use of *many* in this case helps to highlight a point of comparison or contrast, for it shows that the observed correspondence or disparity applies to a great number of people, for instance:

context of a cross-linguistic study of Thai and English quantity words, the similarities and differences in their usage profiles become crucial.

At a more general and theoretical level, I have also identified and presented a number of similarities and differences between the quantity words from perspectives different from previous accounts, i.e., in terms of their collocational, semantic preference and discourse-pragmatic functional profiles. Moreover, the patterns of usage for which I have argued are based directly upon quantitative and qualitative evidence of language use as observed in statistical evidence for collocation on the one hand, and randomly-selected concordance samples on the other. The findings from the present study have highlighted the significance of phraseological and discourse-pragmatic perspectives on quantity words. The study has yielded support to the view that the choice of quantity expressions can perform modal or rhetorical functions (Dubois, 1987; Powell, 1985), and at the same time suggested that the discourse-pragmatic dimension of these quantity words can be realized phraseologically. It has also been demonstrated as well that the two words, though can be used to denote plurality of entities, have different semantic preferences for nouns they quantify. These in turn reflect a contribution from a corpus-based phraseological perspective on Thai-English cross-linguistics research.

It must be noted, however, that the broad semantic property of *lǎay* has been approached in this study through the focus on its two major English equivalents. Future research can approach the use of other English equivalents, especially those denoting the smaller quantity, and such questions as what could possibly prompt translators, interpreters or speakers of Thai to choose particular words to express *lǎay* in English. Also, a cognitive perspective on the various meaning potentials of *lǎay*, i.e., how speakers/readers interpret the scale of quantity expressed by *lǎay* in particular contexts, and on the perception of *lǎay* in comparison with its various English equivalents would help enhance our understanding of *lǎay* in the context of Thai-English cross-linguistic research.

ACKNOWLEDGEMENT

The study reported in this paper is part of the research project *British/Thai Collaboration on National Corpora: Capacity, Mobility, Research*, funded by the British Academy, UK, and Thailand's Office of the Higher Education Commission, under the Newton Fund Mobility Grant Scheme (NG170124). The researcher would like to express her gratitude towards the sponsors, Dr. Andrew Hardie - her UK research host, and two anonymous reviewers of this article.

THE AUTHOR

Raksangob Wijitsopon is an associate professor of English at Department of English, Faculty of Arts, Chulalongkorn University. She is also Head of the Corpus Linguistics for Digital Humanities Research Unit (CLiDH). The areas of her research interests include corpus linguistics, stylistics and discourse analysis.

raksangob.w@chula.ac.th

REFERENCES

- Altenberg, B. (1990). Some functions of the booster. In J. Svartvik (Ed.), *The London-lund corpus of spoken English: Description and research* (pp.193-210). Lund University Press.
- Aroonmanakun, W. (2009). *English-Thai Parallel Corpus*. Chulalongkorn University <http://www.arts.chula.ac.th/~ling/ParaConc/>
- Bach, K. (2000). Quantification, qualification and context: A reply to Stanley and Szabó. *Mind and Language*, 15, 262-283.
- Badia, A. (2009). *Quantifiers in action: Generalized quantification in query, logical and natural languages*. Springer US.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finnegan, E. (1999). *Longman grammar of spoken and written English*. Longman.
- Carden, G. (1976). *English quantifiers: Logical structure and linguistic variation*. Academic Press.
- Channell, J. (1990). *Vague language*. Oxford University Press.
- Deephuengeton, P. (1992). A study of quantifier phrases in Thai. *Kansas Working Papers in Linguistics*, 17(1), 223-255.
- Dichtel, F. (2016). A quantifier used on many occasions: Many evoking diversity in positive sentences. *International Journal of Corpus Linguistics*, 21(1), 80-104.
- Drave, N. (2002). Vaguely speaking: A corpus approach to vague language. In P. Collins, P. Peters & A. Smith, *New frontiers of corpus research* (pp. 25-40). Rodopi.
- Dubois, B. (1987). Something on the order of around forty to forty-four: Imprecise numerical expressions in biomedical slide talks. *Language in Society*, 16, 527-541.
- Gavins, J. (2007). *Text world theory: An introduction*. Edinburgh University Press.
- Hardie, A. (2012). CQPweb - combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17(3), 380-409.
- Hardie, A. (forthcoming). A dual sort-and-filler strategy for statistical analysis of collocation, keywords, and lockwords.
- Hogg, R. (1977). *English quantifier systems*. North-Holland Pub. Co.
- Huddleston, R., & Pullam, G. (2002). *The Cambridge grammar of the English language*. Cambridge University Press.
- Iamworrameth, T. (1995). *Standard English-Thai dictionary*. Ruamsarn.
- Indrambarya, K. (2011). Quantifiers in Thai. *Journal of the Southeast Asian Linguistics Society*, 4(2), 46-60.
- Jaszczolt, K. (2002). *Semantics and pragmatics: Meaning in language and discourse*. Longman.
- Karnden, D., & Wannapok, S. (2009). *Thai-English Dictionary* (4th Edition). Se-Education.
- Keenan, E., & Paperno, D. (Eds.) (2012). *Handbook of quantifiers in natural language*. Springer Netherlands.
- Moxey, M., & Sanford, A. (1993). *Communicating quantities: A psychological perspective*. Lawrence Erlbaum Associates.
- Moxey, M., & Sanford, A. (2000). Communicating quantities: A review of psycholinguistic evidence of how expressions determine perspectives. *Applied Cognitive Psychology*, 14, 237-255.
- Partington, A. (1998). *Patterns and meaning*. John Benjamins.
- Powell, M. (1985). Purposive vagueness: An evaluative dimension of vague quantifying expressions. *Journal of Linguistics*, 21, 31-50.
- Prasithrathsint, A. (2010). Lak Samkhan Nay Kanjamnaek Chanitkham Nay Phasa Thai (Main principles of word classification in Thai). *Selected Articles from the Pan-dialectal Grammar of Thai Symposium on Diversity of Languages in Thai Society, Standard Thai, Thai Regional and Social Dialects: Issues and New Discoveries* (in Thai), ed. by A. Prasithrathsint, S. Rungrojsuwan, and S. Sankaburanurak. ASP.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.

- Ruzaitė, J. (2009). Corpora for applied purposes: A case study of quantifiers in English and Lithuanian. *Eesti Rakenduslingvistika Ühingu aastaraamat. Estonian Papers in Applied Linguistics, 5*, 239-250.
- Savetamalya, S. (1989). *Thai nouns and noun phrases: A lexicase analysis*. [Doctoral dissertation, University of Hawai'i]. SholarSpace. <http://hdl.handle.net/10125/9932>
- Stein, M. (1981). *Quantification in Thai*. [Doctoral dissertation, University of Massachusetts Amherst]. ScholarWorks. <https://scholarworks.umass.edu/dissertations/AAI8110380/>
- Tiangburanatham, V. (2008). *English-Thai & Thai-English dictionary (Completely revised & updated edition)*. Se-Education.
- Tiangburanatham, V. (1995). *A New English-Thai dictionary (Library edition)*. Ruamsarn.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. John Benjamins.