A Comparison of 'little' and 'small' in English and Thai: A Corpus-Based Study

Vilaivan Aroonmanakun[a], Wirote Aroonmanakun[b,*]

[a] Vilaivan.a@litu.tu.ac.th, Language Institute, Thammasat University, Thailand
[b] Wirote.a@chula.ac.th, Department of Linguistics; Corpus Linguistics for Digital Humanities Research Unit, Faculty of Arts, Chulalongkorn University, Thailand
[*] Corresponding author, wirote.a@chula.ac.th

ABSTRACT

This paper studies the two English synonymous words *little* and *small*, and their Thai equivalents nɔ:j4 and lek4. In addition to monolingual and bilingual dictionaries, British National Corpus, Thai National Corpus and an English-Thai parallel corpus were used in this study. It is found that *little* and *small* and their Thai equivalents are not precisely identical in meaning. In addition, there are some meanings that can apply to all these four English-Thai synonymous words. Therefore, a Thai learner of English may not be able to simply equate the use of *little* and *small* with nɔ:j4 and lek4, and vice versa. Moreover, it is also found that *little* and *small* can be translated into Thai with some other words rather than nɔ:j4 and lek4. This indicates the complexity of meaning in which *little* and *small* can connect to other related meanings and be translated into Thai with other words. This study suggests that information provided in dictionaries alone is not sufficient. Learners should learn how to use corpora as a resource to enhance their lexical knowledge in both languages.

Introduction

Learning the similarities and differences of near synonymous words is a crucial and challenging task for language learners. This can be achieved by consulting a dictionary in which different usages of those words are described or by comparing different usages found in a corpus. Though the latter may seem a complicated method, it can successfully yield more informative results. Many studies have been conducted to demonstrate the usefulness of corpora in this field of study. For example, Oh (2000) showed the differences of uses between *actually* and *in fact*. V. Aroonmanakun (2015) compared the differences between *quick* and *fast*. Phoocharoensil (2021) examined the synonyms *persist* and *persevere* by looking through different distribution across genres, collocations and semantic prosody. Lin and Chung (2021) showed the differences between *propose* and *suggest* in academic writing. Lau and Lee (2016) studied two manner adverbs *yiran* and *duanran* 'resolutely' in a Chinese corpus and found the differences in terms of the event structures constructed by the collocated verbs, conjunctions and nouns. Most of the previous studies have focused on analyzing native corpora to identify similarities and differences for foreign language learners because information provided in the dictionaries is not sufficient to understand the differences between near synonymous words. A few studies have compared the use of near synonymous words in both native and learner corpora. For example, Pérez-Paredes and Bueno-Alastuey (2019) used both native and learner spoken English corpora to study three adverbs, *obviously*, *really* and *actually*. Chinese, German and Spanish learners of spoken English corpora are used to revealing different usages from these learners. Hong (2014) used both Chinese corpus and Chinese learner corpus to study the differences between two sets of synonyms, *bian4li4/fang1bian4* 'convenient' and *ren4shi4/zhi1dao4* 'know'. This type of study will add to the differences between learner and native usages of those synonymous words. In this study, in addition to analyzing near synonymous words in English, near synonymous words in the native language of Thai learners were also analyzed. The objective is to reveal similarities and differences not only in the target language but also those in the native language and across languages as well. An English-Thai parallel corpus was also used to verify equivalents of these words in the two languages. This should provide a more ample view for Thai learners to understand synonymous words in English.

Scope and Methodology

This study focuses on two synonymous adjectives, *little* and *small*, and their equivalent translations in Thai, nɔ:j4 and lek4. Three corpora were used as a reference for this study, namely BNC (British National Corpus), TNC (Thai National Corpus), and an English-Thai parallel corpus. BNC and TNC were used to identify the meanings of each word through its concordances and collocates. This should provide a clear picture of how many meanings there are for each word and the similarities or differences of their collocates. BNC online (*Corpora: BNC*) and TNC online (Aroonmanakun et al., 2009) were used in this study. The analysis of *little* and *small* is presented in Section 3, while the analysis of nɔ:j4 and lek4 is presented in Section 4. Section 5 presents the results of the translation of *little* and *small* into Thai. These results will show whether nɔ:j4 and lek4 are equivalent terms to *little* and *small*.

Analysis of *little* and *small*

To explore the differences between *little* and *small*, the Longman Dictionary of Contemporary English Online (Pearson Education Limited, n.d.) was first consulted. The use as a compound or a fixed phrase, e.g., *a little bit, no small degree, small world, small change, small arm*, etc., is more frequent in *small* than in *little*. In this study, however, only the use as a single adjective is considered, and Table 1 contains a comprehensive list. Based on the information from the dictionary, these two adjectives convey eleven different meanings, of which three overlap, i.e. (1) 'small in size', (5) 'young' and (7) 'unimportant'. The rest seem to be specific to either word. In cases where both *little* and *small* can be used, it is not evident whether they are completely interchangeable.

Table 1

Meanings and Examples of little and small from the Longman Dictionary of Contemporary English

Sense	*little* (adj)	*small* (adj)	Meaning
small in size	*a little house*		1 'small in size'
not large in size or amount		*a small car, a small town, a small amount of money*	2 'small in quantity'
something you like or dislike	*a useful little gadget, silly little jokes*		3 'added attitude'

short in time or distance	*a little sleep, a little way*		4 'short in time/distance'
young	*a little boy*	*a small boy*	5 'young'
unimportant	*She gets very angry over little things*		6 'unimportant'
not important		*a small problem, a few small changes*	
slight	*a little smile*		7 'small in degree'
not involving a large amount of money or people		*small business*	8 'small in group' (not involving many people)
letter		*small letters*	9 'small (letter)'
voice		*a small voice*	10 'small (voice)'
look/feel small		*make me look small*	11 'feel small'

Frequency of *little* and *small*

According to the dictionary, both words are ranked in the top 1,000 words in both written and spoken data. However, it is not reported which one is more frequent. To obtain the actual frequency, the BNC was used as a reference corpus. In the BNC, *little* is found 61,932 times compared to *small*'s 42,730 times, which means that *little* is used around 1.4 times more frequently. However, when it comes to use as an adjective merely, *little* appears to have a lower frequency, compared to *small*. *Little* (adj.) is found 30,818 times, while *small* (adj.) is found 41,837 times. The distribution of both words across different genres is shown in Table 2.

Table 2

Distribution of the Adjectives little and small in BNC

	All	Spoken	Fict	Mag	News	non-Acad	Acad	Misc
little	30818	6298	11479	1995	1855	2703	1716	4772
small	41837	2211	8781	3864	2616	7498	6609	10258

Table 2 shows that *small* is more common in formal genres like magazines, newspapers and academic texts, while *little* is preferred in less formal ones like spoken and fiction. This is in accord with what Biber et al. (1996), reported in their work. They found *little* a lot more in conversation than in academic prose, while *small* is found more in academic prose. They explained that this difference is reflected in the preferred collocates of both

words. *Little* tends to co-occur with concrete nouns describing animate beings, commonly found in conversational genre as physical entities that are often mentioned in conversation. *Small*, on the other hand, tends to co-occur with nouns indicating quantity, e.g., amount, number, etc., since measuring quantities and amounts is often found in academic prose.

To verify the meanings concluded from the dictionary, 100 concordances of each word were randomly sampled and analyzed in this study. The result confirms the meanings found in Table 1, with some additional meanings for both *little* and *small*. *Little* is found with the meaning of 'short in length' (4), e.g., *little poem, little hand-sawing*. This meaning is similar to 'short in distance' (4) because it is related to a space domain. In addition to the 'unimportant' meaning (6), *little* is used to describe the quality of 'being not complicated' (12), e.g., *little analysis, little job, little work*. Other additional meanings are 'small in group' (8), e.g., *little troupe*; 'small (voice)' (10), e.g., *little voice*; 'low in status' (13), e.g., *little parson, little muttering peasant*.

With respect to *small*, the additional meanings are as follows: 'small in quantity' (2) e.g., *a small number of …, a small proportion of …, a small selection*; 'short in time/distance/length' (4), e.g., *a few small hours, small time, a small jump*; and 'small in degree' (7), e.g., *a small difference, small experience*. In the case of the overlapped meanings with *little*, the collocates of both words found in these samples are not similar. For example, when used to denote 'small in quantity' (2), *little* collocates with *chance, agreement, money, progress, value*, while *small* collocates with *amount, number, proportion, selection*. It seems that *small* can be used for 'small in quantity' only when quantified nouns are added. Even when *little* and *small* can be used to indicate 'small in size' (1), they do not share the noun collocates in these samples. Only *room* and *town* are found to co-occur with both words. Anyway, since this analysis is based only on 100 samples for each word, a definite conclusion still cannot be drawn from these findings. To explore the similarities or differences of the collocates, all noun collocates are analyzed in the next section.

Table 3

The Number of Meanings Analyzed from 100 Samples

	Meaning	*little*	*small*
1	'small in size'	30	49
2	'small in quantity'	25	17
3	'added attitude'	6	-
4	'short in time/distance/length'	7	4
5	'young'	11	2
6	'unimportant'	1	-
7	'small in degree'	8	6

8	'small in group'	1	19
9	'small (letter)'	-	2
10	'small (voice)'	2	-
11	'feel small'	-	-
12	'not complicate'	3	-
13	'low in status'	2	-

Collocates of *little* and *small*

To confirm whether the uses of *little* and *small* are as described in the dictionary and in Biber et al. (1996), top noun collocates of the two words were extracted from the BNC. First, we extracted the top collocates based on statistical values, which is mutual information (MI). This method will show the collocates that are statistically significant with *little* or *small*. However, their collocates tend to be among the low frequency words that occur exclusively with either *little* or *small*. The collocates found by this method then are not ordinary words. For example, the top collocates of *little* are *schoolbook, piggies, weirwold, dorrit, blighters, emme, etc.,* and the top collocates of *small* are *intestine, mammal, cetaceans, mercies, resection, wader, etc.* With this method, only 3 collocates, *clique, cubicle* and *shiver*, co-occur with both words. This result suggests that using MI does not yield collocates that are commonly used.

Since no other statistical method was provided for using the BNC online, the top 100 noun collocates were extracted based on their frequency. By this method, 33 collocates are found to co-occur with the two words. These are *boat, box, boy, boys, car, child, children, cottage, fish, garden, girl, group, hole, house, island, man, office, ones, party, piece, pieces, place, room, shop, smile, table, thing, things, town, village, voice, way,* and *world*. In other words, about one-third of the collocates are shared by the two words. This is because these collocates are high frequency words. The chances that they would co-occur with both *little* and *small* will, accordingly, be high. However, the differences in uses of the above-mentioned shared collocates are still in question.

Differences between *little* and *small*

To explore the differences in uses of the shared collocates, up to 100 sample concordances of five collocates, namely *boat, fish, boy, town* and *group*, were extracted and analyzed. These five words were selected to represent different collocate types respectively, i.e., a concrete noun (nouns referring to inanimate physical objects), an animate noun (nouns referring to animate beings, except humans), a human (nouns referring to human beings), a place (nouns referring to places or locations) and a quantitative noun (nouns referring to collectives). The distribution of these collocates across genres is shown in Table 4. Up to one hundred samples of each collocate were

extracted and analyzed. Only three collocates, *little boat*, *little fish* and *little boy* were found fewer than 100. The number in parenthesis is the total occurrence of *little/small* and their collocates. The color is used to highlight items that seem to be different.

Table 4

Distribution of Collocates of Little and Small Across Genres

		All	Spoken	Fict	Mag	News	non-Acad	Acad	Misc
boat	*little*	(46)46	15%	70%	2%	2%	0%	0%	11%
	small	(102)100	11%	38%	8%	4%	16%	0%	23%
fish	*little*	(39)39	8%	31%	26%	25%	13%	3%	15%
	small	(146)100	0%	10%	34%	7%	13%	4%	32%
boy	*little*	(98)98	23%	1%	0%	31%	0%	10%	35%
	small	(226)100	7%	1%	0%	24%	7%	4%	57%
town	*little*	(150)100	9%	21%	0%	11%	20%	0%	39%
	small	(406)100	5%	1%	0%	18%	0%	7%	69%
group	*little*	(101)100	14%	52%	3%	2%	12%	3%	14%
	small	(761)100	7%	0%	0%	1%	0%	81%	11%

From Table 4, it is still unclear how differentiate between the uses of *little* and *small*. Considering the overall uses, though, *small* is preferred in formal genres while *little* is more common in informal genres (see Table 2). However, this finding is not repeated in all the five collocates here. For *boat* and *fish*, *little* is found to co-occur with less frequency than *small* in most genres due to its fewer occurrences in total use. For *town*, *little* seems to be preferred in both fiction and non-academic texts but not in academic texts. For *group*, *little* occurs more in fiction while *small* is preferred in academic texts. But for *boy*, *little* seems to be preferred in spoken texts, though its total use is less than that of *small*. It is possible that *little boy* is used with a sense of added attitude towards the subject in spoken texts. Nevertheless, these differences are minor and do not clearly indicate the differences between the two words.

Since *little* has one meaning that is related to the feelings of like and dislike, it is likely that *little* in a good sense would appear in a positive context. With this assumption, a sentiment analysis program, in which a positive or negative sentiment is determined from words in the context, was applied. The concordances of *little/small* and the five collocates were analyzed for their sentiment degree. A Python library (Hutto & Gilbert, 2014) was used to analyze each concordance line. The sentiment score ranging from -1 to 1 is calculated from the context of each concordance line. In this study, a positive sentiment is marked when the score is greater than 0.3; a negative is marked when the score is less than -0.3; otherwise, a neutral sentiment is marked. (See

the Python program for sentiment analysis in Appendix A). Some examples of sentiment analysis results are shown below:

Table 5

Examples of Sentiment Analysis

Context	Sentiment
thought about it day and night. My master liked to go fishing in a **little boat**, and he always took me with him. A man called Moely,	*pos*
as she was a bit claustrophobic, stepping over the side into that ridiculous **little** rubber **boat** was terrifying. A fair swell was running and the boat bobbed	*neg*
again. Down, and up again. And it was coming nearer! A **little boat** with two people in it. He held the rock and stared at it	*neu*
going on what (unclear) a great big boat and it was in fact only a **small** water **boat**. And we were all enjoying it and looking forward to going too	*pos*
in Hungary outside Budapest. Two leading liberals had drowned. There were rumours the **small boat** had been hit by a much larger vessel. The second picture showed what	*neg*
and pull it into the small boat? (SP:PS22D) Pull it into the **small boat**, lay it on a box and then we'd pull the chain in	*neu*
our big winner at the Games but the biggest winner of all was his **little boy**, Edward. He is a champion on his super bike. A world	*pos*
Cameron plays with the big boys but underneath his brash exterior there lurks a **little boy**, haunted by an event in his childhood. Between the sheets with his	*neg*
the children are squawking, # chasing each other, picking fruit; # the **little boy** brings me a flower and a carrot. # We're drinking must --	*neu*
eye in the... # HEART-THROB Patrick Swayze gets one in the eye from a **small boy** in the forthcoming heat-and-dust saga City Of Joy. As a doctor who founds	*pos*
child called Kathy listened as her mother screamed at her husband who had killed a **small boy** in one such accident. Again this was shown after 9pm. Last year	*neg*
speed. It began to rain. In the porch of a wooden shack a **small boy** dangled a land crab on the end of a makeshift fishing rod; it	*neu*

Table 6

Sentiment Analysis of Little and Small

		All	Negative	Neutral	Positive
boat	*little*	46	24%	41%	35%
	small	100	20%	47%	33%

fish	*little*	39	8%	31%	62%
	small	100	16%	46%	38%
boy	*little*	98	26%	38%	37%
	small	100	19%	40%	41%
town	*little*	100	14%	45%	41%
	small	100	13%	51%	36%
group	*little*	100	13%	54%	33%
	small	100	11%	45%	44%

In Table 6, the distinction between positive and negative uses of the two words is not clearly noticeable. Positive and negative sentiments can be found in both *little* and *small*, though negative senses seem to be less used. It can be concluded that for the noun collocates that can be used with either *little* or *small*, there are no rigid rules governing their use. Some of them are interchangeable in the same context. However, this does not indicate that there is no preference for use in all collocates. There are several collocates that prefer *little* over *small* and vice versa. When using the function "compare" between *little* and *small* and assigning NOUN as a collocate, the ratio of collocates between these two words is used to calculate collocational scores and the top collocates of each word are listed.

Table 7

Noun Collocates that Prefer Little Over Small

Word	*little*	*small*	*little/small*	Score
while	433	0	866	597.5
evidence	398	1	398	274.6
impact	158	0	316	218
sense	112	0	224	154.5
progress	105	0	210	144.9
knowledge	90	0	180	124.2
thought	84	0	168	115.9
reason	80	0	160	110.4
attention	288	2	144	99.4
sympathy	68	0	136	93.8
relevance	60	0	120	82.8
bastard	57	0	114	78.7
icing	55	0	110	75.9
bitch	53	0	106	73.1
bit	1725	17	101.5	70
faith	49	0	98	67.6

attempt	43	0	86	59.3
contact	43	0	86	59.3
incentive	43	0	86	59.3
sleep	43	0	86	59.3

Table 8

Noun Collocates that Prefer Small Over Little

Word	*small*	*little*	*small/little*	Score
proportion	324	0	648	939.2
minority	204	0	408	591.3
bowel	195	0	390	565.3
intestine	160	0	320	463.8
hours	147	0	294	426.1
quantities	136	0	272	394.2
scale	258	1	258	373.9
percentage	115	0	230	333.4
mammal	105	0	210	304.4
mammals	101	0	202	292.8
schools	69	0	138	200
investors	64	0	128	185.5
amounts	244	2	122	176.8
businesses	482	4	120.5	174.7
farmers	119	1	119	172.5
claims	57	0	114	165.2
quantity	57	0	114	165.2
firms	414	4	103.5	150
businessmen	50	0	100	144.9
sector	47	0	94	136.2

It is clearly evident from Tables 6 and 7 that several noun-collocates tend to co-occur with either *little* or *small* only. These collocates need to be emphasized when teaching the distinction between these two adjectives. They need to be emphasized for learners for a clear understanding of the distinction between these two adjectives.

Analysis of *nɔ:j4* and *lek4*

For Thai learners, there are two Thai synonymous words, nɔ:j4 and lek4 , which are identical in meaning to *little* and *small*. Unlike English dictionaries, most Thai dictionaries describe different senses of each entry briefly. For example, the Thai Royal Institute dictionary (2011) describes the meanings of lek4 as, first, (one thing) being smaller compared to another one;

then not being of a large/big size and, finally, being unimportant. As for nɔ:j4, two entries are listed. The first entry has two senses which are the opposite of ma:k3 'many' and the opposite of jaj2 'big'. The first sense is about the amount of something while the second is about the status since the example of a lower ranking officer is given. The second entry has two senses. The first sense describes the lower status of a person, while the second sense denotes an attitude towards something as adorable. In addition, lek4 and nɔ:j4 can be combined as a compound lek4 nɔ :j4 , which means 'a little', and 'not important'. Based on these definitions, nɔ:j4 and lek4 seem to be distinct from each other, while nɔ :j4 and lek4nɔ:j4 have one overlapping meaning, i.e., 'unimportant'. These descriptions are summarized in Table 9, in which the corresponding meanings of *little/small* in Table 3 are marked in the right-hand column.

Table 9

Meanings of lek4, nɔ:j4 and lek4nɔ:j4 from a Thai Dictionary

Sense	lek4	nɔ:j4	lek4nɔ:j4	Meaning
smaller in size	*X* lek4kwa:2 *Y* 'X is smaller than Y'			1 'small in size'
not big	ba:n3laŋ5ni:4 lek4 'this house is small'			1 'small in size'
unimportant	rɯ:aŋ3lek4 'small matter'		rɯ:aŋ3 lek4nɔ:j4 'little matter'	6 'unimportant'
not many		fon5nɔ:j4 'little rain', mi:1ŋɤn1 nɔ:j4 'have little money'		2 'small in quantity'
low in rank		kʰa:3.ra:t3.ka:n1cʰ an4pʰu:3nɔ:j4 'low rank officer'		13 'low in status'
less important		pʰu:3nɔ:j4 'inferior'		13 'low in status'
feeling like		dek2nɔ:j4 'little boy'		3 'added attitude'

Frequency of nɔ:j4 and lek4

For a better understanding of the two Thai synonymous words nɔ:j4 and lek4, the Thai National Corpus (TNC) is used as a reference corpus. The Thai National Corpus web[1] and CQPweb at Lancaster[2] are used to query the corpus. Of the total findings, lek4 has 18,993 occurrences, while nɔ:j4 has 29,422 occurrences, and lek4 nɔ:j4 has 3,997 occurrences. To investigate the actual use of these words, 100 samples of concordances for each word, lek4, nɔ:j4 and lek4 nɔ:j4, were randomly extracted and analyzed. Eight additional meanings are found in these samples. They are highlighted in Table 10. The number at the head of each row corresponds to the meaning number of *little/small* in Table 3. The word lek4 can refer to 'small in group' (8), (e.g., cʰum1 cʰon1 -lek4 'community-small') 'young' (5) and 'small (voice)' (10) in a similar way to *small*. But for the meaning 'not complicated' (12), lek4 corresponds to *little*, not *small*. In Thai, a research project is considered 'not complicated' by using the word kʰa2 na:t2 -lek4 'size-small'. The words nɔ:j4 and lek4 nɔ:j4 can refer to 'small in degree' (7), e.g., ʔo:1 ka:t2 -nɔ:j4 'chance-little', lot4 loŋ1 -lek4 nɔ:j4 'decrease-little_bit', and 'short in time' (4), e.g., we:1 la:1 -nɔ:j4 'time-little', du:1 -nan3 -du:1 -ni:3 -lek4 nɔ:j4 'look-those-look-these-little_bit', which is similar to the use of *little*. As for lek4 nɔ:j4, its usages seem to correspond to nɔ:j4 rather than lek4. It can be used for 'small in quantity' (2), 'short in time' (4), 'small in degree' (7) like those in nɔ:j4. lek4 nɔ:j4 is a type of compound in Thai, called 'semantic doublet' (Vongvipanond, 1985). Besides semantic doublets, reduplication is another type of compound often used in Thai to intensify the meaning. Therefore, it is not surprising to find reduplications of the three words, lek4 lek4 'small-small', nɔ:j4 nɔ:j4 'little-little', and lek4 lek4 nɔ:j4 nɔ:j4 'little_bit-little_bit' in these samples. Moreover, another type of compound, appositive compound (Booij, 2007:98), e.g., ma:k3 nɔ:j4 'many-little' is also found to denote the meaning of 'more or less' and another semantic doublet _lek4 _nɔ:j4 is also found in the samples.

Table 10

Meaning Analysis of lek4, nɔ:j4, lek4nɔ:j4

	Meaning	lek4	nɔ:j4	lek4 nɔ:j4	lek4 lek4	nɔ:j4 nɔ:j4	lek4 lek4 nɔ:j4 nɔ:j4	_ lek4 _nɔ:j4
1	'small in size'	34			13			
2	'small in quantity'		45	27				
3	'added attitude'		7			1		
4	'short in time/ distance/ length'		2	23				
5	'young'	15			5			
6	'unimportant'	2		12	1		8	3
7	'small in degree'		6	38				
8	'small in group'	13			1			
10	'small (voice)'	1						
12	'not complicated'	1			1			

Though the meaning 'low in status' (13) is found in the samples, it is analyzed as a compound, e.g., mi:a1 nɔ :j4 'concubine', cʰon1 klum2 nɔ:j4 'minority group'. Thus, it is not presented in the above table. In fact, the examples presented in the Thai dictionary, kʰru:1 nɔ :j4 'lower rank teacher', pʰu:3 nɔ :j4 'lower rank person', can also be analyzed as a compound. Therefore, this meaning is omitted in Table 10. The total meanings found for lek4 and nɔ:j4, then, are ten meanings.

Collocates of lek4, nɔ:j4 and lek4nɔ:j4

The analysis in the previous section suggests that lek4 and nɔ :j4 are used in different contexts. lek4nɔ:j4 can be used in the same meanings, mostly as nɔ:j4 because it is a semantic doublet, which is an intensifier of nɔ:j4. However, analyzing their collocates will be required to verify this conclusion.

Since TNC does not have POS tags and the collocates of these words can appear in many positions on the left side of the words, CQPweb was applied to limit the search for only the top collocates that are 1-5 words on the left side of lek4, nɔːj4, and lek4nɔːj4 and to rank the results by Log Ratio values (Hardie, 2014). The result should yield many candidates but not all of them are true collocates. To confirm that the words on the collocation lists are the true collocates, the researchers need to manually look through the concordance lines of each result.

To manually analyze the three lists of collocates, the top 100 collocates extracted from each list were reviewed and marked for the one that is a true collocate. The number of collocates found for lek4, nɔːj4 and lek4nɔːj4 are 21, 28, and 28 respectively. Collocates shared by two words are shown in Table 11.

Table 11

Top Shared Collocates of lek4, nɔːj4, and lek4nɔːj4

collocate	lek4	nɔːj4	lek4nɔːj4	notes
dek2 'child'	'young'	'added attitude'		
ʔoŋ1 'person'	'young'	'added attitude'		
lem3 CLSS	'small in size'	'added attitude'		
kuŋ3 'shrimp'	'small in size'	'added attitude'		
sat2 'animal'	'small in size'	doublet		sat2lek4sat2nɔːj4, sat2nɔːj4jaj2
tum2 'pimple'	'small in size' doublet	doublet		tum2lek4tum2nɔːj4
raːn4 'store'	'small in size' doublet	doublet		raːn4lek4raːn4nɔːj4, raːn4lek4lek4
hɔːŋ3 'room'	'small in size' doublet	doublet		hɔːŋ3lek4hɔːŋ3nɔːj4
jaj2 'large'	compound	compound		jaj2lek4, jaj2nɔːj4
rɯːaŋ3 'matter'	'unimportant'		'unimportant'	
puːaj2 'sick'		'unimportant'	'unimportant'	
jim4 'smile'		'small in degree'	'small in degree'	jim4nɔːj4 nɔːj4

sa2.wa:ŋ2 'bright'	'small in degree'	'small in degree'
kʰo:ŋ4 'curve'	'small in degree'	'small in degree'
kʰɛŋ5 'hard'	'small in degree'	'small in degree'
pʰɤ:m3 'increase'	'small in quantity'	'small in quantity'

Both lek4 and nɔ:j4 are found to denote the meanings of 'small in size' and 'young'. Only nɔ:j4, however, can be used with an added attitude, i.e., the implied feeling of like or love towards an object referred to by the noun collocates. To intensify the meaning of 'small in size', a reduplication lek4lek4 is used. Another kind of semantic doublet, _lek4_nɔ:j4, is used to indicate plurality. This kind of semantic doublet is productive and found in the samples, e.g., tum2lek4tum2nɔ:j4, ra:n4lek4ra:n4nɔ:j4, hɔ:ŋ3lek4hɔ:ŋ3nɔ:j4, etc.

Besides semantic doublets, another type of compound, appositive compounds, jaj2lek4, jaj2nɔ:j4, nɔ:j4jaj2, lek4jaj2, ma:k3nɔ:j4, are also found. These are compounds formed by either of the two words, lek4 or nɔ:j4, being combined with its oppositions, jaj2 or ma:k3. To signify the meaning 'big or small' or 'more or less'. The opposition word can be placed before or after the word. Only one combination, nɔ:j4ma:k3, is not possible because it means 'very little' due to the word order and the meaning of ma:k3 'many'. (cʰɔ:p3-ma:k3 'like-more' would mean 'like (something) very much'.)

One remarkable use of nɔ:j4 is that it is often found collocated with a negation in Thai for the meaning of 'many'. That is, maj3nɔ:j4 is found 4,460 times (128.2/million), compared to maj3lek4, which is found only 465 times (13.4/million).

In addition to the shared collocates above, lek4 and nɔ:j4 seem to have different sets of preferred collocates as well. These collocates are listed in Table 12. Most of the collocates of lek4 are nouns and the meaning 'small in size' is applied. As for nɔ:j4, most of the collocates are predicates. Though the preference lists of nɔ:j4 and lek4nɔ:j4 in Table 12 are distinct from each other, collocates that are a predicate can co-occur with either nɔ:j4 or lek4nɔ:j4. This is because lek4nɔ:j4 is usually an intensifier of nɔ:j4.

Table 12

Preferred collocates of lek4, nɔ:j4, and lek4nɔ:j4

Word	Collocates
lek4	ra:ŋ3 klum2 nɔ:ŋ4 hɔ:ŋ3 ra:n4 met4 sat2 tum2 kɛ:w3 cʰɔ:ŋ3 run3 hun2 kʰem5 pa:j3 cʰɔ:3 kriŋ2 pʰum3
nɔ:j4	pen1 hen5 num2 kʰa:ŋ3 ca:j2 sa:ŋ3 pʰa:n2 tʰuk4 cʰɯ:a3 cʰa:ŋ4 nuŋ3 tʰuŋ3 sɔ:m4 tʰa:j2
lek4nɔ:j4	ta:ŋ2 kep2 cʰu:aj3 ri:2 pli:an2 rew1 jɯ:aŋ4 klin2 bi:aŋ2 kʰem1 da:ŋ2 kʰem3 da:j3 kɔ:n2 ʔɔ:n2 pri:aw3 bɯ:a2 bɛŋ2 ʔun2 tʰiŋ4 li:aw4

The analysis of collocation confirms the ten meanings found for the word lek4, nɔ:j4, and lek4nɔ:j4. Lek4 is used for the meaning of 'small in size' while nɔ:j4 is used for that of small amount or degree of something. When nɔ:j4 is used to signify 'small in size', it is often used with added attitude toward the object.

Comparison of English and Thai near synonyms

The analyses in Sections 3 and 4 seem to suggest that *little/small* and nɔ:j4/lek4 do have similar meanings. Oxford-River Books Concise English-Thai Dictionary (Mallikamas et al., 2010) gives equivalent translations for *little* (adj.) as (a) lek4 'small'; (b) jaw1waj1 'young'; (c) ti:a3 'short (person)' or san3 'short (time or distance)'; (d) lek4nɔ:j4 'not much'; (e) lek4, lek4nɔ:j4, nɔ:j4 'trivial'. As for small (adj.), this dictionary gives equivalent translations as (a) lek4 'small in size', kʰap4kʰɛ:p3 'small (area)', kʰɔ:j3 'not loud'; (b) kʰa2na:t2lek4 'small in scale (business)'; (c) jaw1 'young'; (d) kʰa2na:t2lek4 'small letter'; (e) nɔ:j4 'not much'; (f) mi:1 kʰa:3nɔ:j4 'trifling'; (g) rɔ:ŋ1 'minor'; (h) kʰɛ:p3 'petty'; (i) la4ʔi:at2 'fine'. It can be seen from this bilingual dictionary that there are other Thai translation terms for *little/small* apart from nɔ:j4 and lek4.

Translation of little and small into Thai

To find out more about Thai equivalent terms for translating *little/small*, an English-Thai parallel corpus was used. Five hundred samples of translation containing the word *little/small* were randomly extracted from the English-Thai parallel concordance[3]. Its equivalent term in Thai was identified from each translation and summarized in Table 13. Analysis of

meanings here is based on Thai translation words not *little* or *small* in English. Although each sense of *little* or *small* in the source text can be identified with the meanings listed in Table 3, the analysis of Thai translated words should provide the actual meanings of nɔːj4 or lek4 and other related words.

Table 13

Translations of little and small

Word	Translation	Meaning
little	lek4nɔːj4 30, nɔːj2 28, nit4nɔːj2 22, nit4nɯŋ1 10, nit4nɯŋ2 4, nit4nit4, nit4 3, nɔːj4ma:k3 2, nɔːj2nɔːj2 2, mɛː4pʰiːaŋ1nɔːj4nit4 1, tʰiː1.la4nɔːj4nɔːj4 1, nɔːj4 1, nit4diːaw1 1, nɔːj2nɯŋ1 1	(7) 'small in degree'
	lek4 37, lek4lek4 35, lek4nit4diːaw1 2, nit4diːaw1 1, nɔːj2nɯŋ1 1, lek4lek4baːŋ1baːŋ1 1, tuːa1lek4tuːa1nɔːj4 1	(1) 'small in size'
	nit4nɔːj2 14, nɔːj2 14, nit4nɯŋ1 8, nit4diːaw1 6, nɔːj4 3, nɔːj4nɔːj2 3, lek4lek4nɔːj4nɔːj4 2, lek4nɔːj4 2, sak2nɔːj2 2, sak2nit4 1, ʔiːk2nɔːj2 1	(2) 'small in quantity'
	nɔːj2 23, nit4nɔːj2 8, nit4nɯŋ1 3, nɔːj4 2, ʔiːk2nɔːj2 2, sak2nɔːj2 1, sak2nit4 1, pɛːp4 1, pɛːp4nɯŋ1 1, ʔiːk2pɛːp4 1, san3san3 1	(4) 'short in time/ distance/length'
	nit4nɔːj2 8, lek4lek4nɔːj4 nɔːj4 3, lek4 lek4 2, lek4nɔːj4 1, nit4nit4nɔːj2nɔːj2 1	(6) 'unimportant'
	nɔːj4 9, nɔːj4nɔːj4 4	(3) 'added attitude'
	lek4 4, khon1lek4 1	(5) 'young'
	lek4lek4 1	(8) 'small in group'
	baw1baw1 4	(10) 'small (voice)'
	tam2tɔːj3 1	(13) 'low in status'
	maj3kʰɔːj3 4, maj3tʰaw3raj2 1, maj3daːj3maːk3 1, maj3daːj3maːk3nak4, maj3n̂iː 1, maj3naːn1 1, maj3maːk3 1, maj3maːk3nak4 1	negation
	li4tɤn3 12 (*little*)	transliteration
	kʰɔːn3kʰaːŋ3 5, baːŋ3 3, pʰiːaŋ1 2, paj1 2, paj1nɔːj2 2, kɤːn1paj1 1, kɤːn1paj1nɔːj2 1, pʰiːaŋ1kʰɛː3 1, kʰɛː3 1, kʰraːw3 kʰraːw3 1, ŋuː1 ŋuː1 plaː1 plaː1 1	others
small	lek4 319, lek4lek4 51, lek4cɔːj3 2, lek4sut2 2, lek4nit4nɔːj2 1, lek4nɔːj2 1, siːaw3 1, nit4diːaw1 1, bɔːp2baːŋ1 1, tʰaːt2lek4 1	(1) 'small in size'

lek4lek4 8, nit4nɔːj2 4, lek4 lek4nɔːj4 nɔːj4 4, lek4nɔːj4 2, nit4nɯŋ1 2, lek4 2	(6) 'unimportant'
lek4nɔːj4 6, nit4nɔːj2 4, nɔːj4 2, nit4nɯŋ1 2, nit4diːaw1 1, nɔːj4nɔːj4 1	(2) 'small in quantity'
nit4nɔːj2 2, lek4lek4 1, lek4nɔːj4 1, caːŋ1caːŋ1 1	(7) 'small in degree'
lek4lek4 4	(8) 'small in group'
lek4 1, baw1baw1 1, baw1 1	(10) 'small (voice)'
lek4 1	(5) 'young'
nit4nɔːj2 1, tiːa3tiːa3 1	(4) 'short in time/ distance/length'
lek4 1	(9) 'small (letter)'
raj4 kʰaː3 2, raj4 kʰwaːm1maːj5 1	(12) 'low in status'
sa2mɔːn1 2 (*small*)	transliteration
maj3jaj2maːk3 2, maj3kiː2kʰon1 1, maj3rap4ruː4 1	negation
*lek4 1, jaj2nɔːj4 1, *kʰa2naːt2lek4 1, tʰiːʒriːap3ŋaːj3 1, baːŋ1(weːɪ.laːɪ) 1, la4ʔiːat2 1	others

It can be seen that *little/small* are not always translated into nɔːj4, lek4, or lek4nɔːj4. Variant forms of these words are found in translation such as, lek4nit4diːaw1, lek4nɔːj2, nɔːj4nɔːj2, etc. *Small* seems to be translated mostly into lek4 (including its variants), while *little* seems to be translated into various forms. They can be omitted in the translation in some cases. Twenty six percent of *little* (130/500) is omitted in the translating process, while only ten percent of *small* (50/500) is omitted. There are words other than nɔːj4/lek4 being used in the translation. Many of them have similar meanings to nɔːj4/lek4/lek4nɔːj4. For example, nit4nɔːj2, nit4diːaw1, nit4nɯŋ1, etc., mean 'small in degree' or 'small in quantity'; pɛːp4, pɛːp4nɯŋ1 mean 'short in time'; tam2tɔːj3 means 'low in status'; kʰɔːj3, baw1 mean 'small (voice)'; tiːa3tiːa3 means 'short (height)' or san3san3 means 'short (distance)'. In many cases, since the translation does not need to be done word for word, *little/small* are often omitted or translated using other words to keep the overall meaning of the text. We found other words in the translation of *little/small*, such as kʰɔːn3kʰaːŋ3, kʰɛː3, pʰiːaŋ1, kɤːn1paj1, baːŋ3, kʰraːw3kʰraːw3, etc.

Comparing *little/small* and nɔːj4/lek4/lek4nɔːj4

According to the above analysis, *little/small* can be translated into nɔːj4, lek4, lek4nɔːj4 including their variant forms and other related words. Some of the variant forms are more favored than these target words in certain

meanings. For example, all variant forms of nɔ:j2 and nit4 are more used in the senses of 'small in degree' and 'small in quantity' than those of nɔ:j4/lek4nɔ:j4. Some senses of *small*, such as 'small (voice)', 'short in distance' are translated into other Thai words 'baw1' and 'ti:a3'. In addition, *little/small* can have their meanings extended into other related senses through being translated into other words. These words usually have other English equivalent terms apart from *little/small*. This explains why the bilingual English-Thai dictionary gives more meanings than those defined in English dictionaries, i.e., kʰap4kʰɛ:p3 'small (area)', la4ʔi:at2 'fine'. In a Thai-English translation text, it might be possible to find lek4 and nɔ:j4 being translated into words other than *little* or *small* as well. In fact, a list of words sharing similar meanings to *little/small* can be obtained from a thesaurus, e.g., *low, slight, minor, compact, tiny, minute, short*, etc. Therefore, this study will focus on a comparison between *little/small* and lek4/nɔ:j4/lek4nɔ:j4 to explore whether they are equivalent terms or not.

Table 14

Mapping between little/small and nɔ:j4/lek4

		little	nɔ:j4	lek4nɔ:j4	small	lek4
1	'small in size'	X			X	T C
2	'small in quantity'	X	T C	T C	X	
3	'added attitude'	X	T C		-	
4	'short in time/ distance/length'	X	T C	C	X	
5	'young'	X			X	T C
6	'unimportant'	X		T C	-	T C
7	'small in degree'	X	C	T C	X	T
8	'small in group'	X			X	T C
9	'small (letter)'	-			X	T C
10	'small (voice)'	X			-	T C
11	'feel small'	-			-	
12	'not complicated'	X			-	C
13	'low in status'	X			-	

Table 14 shows the meanings of *little/small* from Table 3 and those of nɔ:j4/ lek4/lek4nɔ:j4 found from the analysis of TNC (marked by C) and Thai translation (marked by T). It can be seen that nɔ:j4/lek4/lek4nɔ:j4 are not exactly equivalent in meaning to *little/small*. While *little* and *small* can be used for 'small in size', 'young' or 'small in group' only lek4 can be used in these

senses. *Little* seems to be used in more varied senses than others. Some senses, such as 'short in length', 'low in status' are translated into other Thai words, not lek4/nɔ:j4/lek4nɔ:j4. In Thai, only lek4 can be used for 'small in size', while nɔ:j4/lek4nɔ:j4 seem to be used for 'small in quantity'. This is different from *little*/*small* because *little*/*small* can be used in both senses. Given that lek4nɔ:j4 is an intensifier of nɔ:j4, those senses covered by both nɔ:j4 and lek4nɔ:j4 will be close to *little* than *small*. nɔ:j4/lek4nɔ:j4 can be used like *little* for the meanings of 'small in quantity', 'added attitude', 'short in time/distance/ length', 'unimportant' and 'small in degree', while lek4 can be used like *small* for the meanings of 'small in size', 'young', 'unimportant', 'small in degree', 'small in group' and 'small (letter)'. However, some senses of *little*, i.e., 'small in group', 'small (voice)' and 'not complicated' can equate with that of lek4 but not nɔ:j4. Moreover, some senses of *little* are mapped to other Thai words, such as 'low in status' - tam2tɔ:j3, 'small (voice)' - baw1baw1. This result indicates that the two synonymous words *little*/*small* and nɔ:j4/lek4 are not exactly equal, which, consequently, can confuse language learners.

Discussion

In practice, learners can use their native language knowledge to equate the target word in the other language. Thai learners might try to equate nɔ:j4 with *little* and lek4 with *small*, but the analysis here shows that they are not truly equivalent in all contexts. Learners should be aware of this fact. They need to understand the senses and usages of the target word by first consulting a learner dictionary. All senses and patterns of use of the target word, usually compiled from a corpus, are listed in the dictionary. But if information provided in the learner dictionary is not sufficient, learners can use a large corpus to search for the actual uses themselves. In cases where a comprehensive dictionary does not yet exist in the target language, corpora are greatly needed to derive different senses and usages of the target word. In addition, when using a large corpus, collocates that are preferred for each synonymous word can be easily drawn. Learners can use this finding to confirm their word choice but when the same collocates are found for both synonymous words, whether or not they are truly interchangeable needs to be examined.

Learners can learn more about synonymous words by using a bilingual dictionary to look for equivalent words in the target language. They will usually get a list of possible translated words but this information alone does

not help them in selecting the right target word. They still need to consult a learner dictionary or a corpus to understand the correct uses. Using a parallel corpus can confirm that it is possible to use other target words. In this study, a sentence containing *little/small* may not have nɔːj4/lek4 in the translation. Other words related to some senses of *little/small* can be seen in the parallel corpus. Learners, then, should be aware of other word choices. They should learn how to use words not limited solely to *little/small*.

In sum, *Little/small* are frequently used words describing the small size and quantity of objects. Since these are common concepts shared among languages, it is easy to find 'equivalent' words in other languages. However, this study shows that not all the senses used in one form can be mapped on to the 'equivalent' word in another language. Language learners must learn which senses can equate with that 'equivalent' word and which senses cannot. In addition, these words, when translated into other languages, can have their meanings extended into related senses through other word forms. This indicates the complexity of meanings which can be expressed by a variety of related words. When learning a foreign language, learners, therefore, must be aware of the complexity of meanings across languages. Equivalent terms in bilingual dictionaries and learner dictionaries alone can provide only basic usages of the words. Learning to use and acquire more information from native and parallel corpora is useful for language learners truly to understand the language.

Moreover, in cases of words that are culturally bound, whose meanings do not precisely equate with words in another language, this may result in even more difficulties for language learners, as the information provided in bilingual dictionaries cannot be exactly mapped. For example, cooking terms like *grill, roast* and *toast*, are translated into Thai as *grill* – (1) piŋ3 or jaːŋ3, *roast* – (1) piŋ3 or jaːŋ3; (2) ʔop2, *toast* – (1) piŋ3 (Oxford-River Books English-Thai Dictionary). This might cause some difficulties for learners since the translations in the target language overlap one another and do not truly show the exact meanings of these three words, *grill/roast/toast*. Though learner dictionaries may provide some explanation of these three terms, the meanings of cooking terms in Thai may not exactly match. Thai learners of English must learn not only the meanings in English but also the meaning in Thai as well. Therefore, it is advisable that they learn how to use corpora as a resource to enhance their learning of all foreign languages.

Acknowledgements

About the Authors

Vilaivan Aroonmanakun: An Assistant Professor in the Language Institute, Thammasat University. Her research interests are language studies and corpus linguistics.
Wirote Aroonmanakun: An Associate Professor in the Department of Linguistics; Corpus Linguistics for Digital Humanities Research Unit, Faculty of Arts, Chulalongkorn University. His research interests are corpus linguistics and computational linguistics.

Endnotes

[1] https://www.arts.chula.ac.th/ling/tnc3/
[2] https://cqpweb.lancs.ac.uk/
[3] https://www.arts.chula.ac.th/~ling/ParaConc/

References

Aroonmanakun, V. (2015). *Quick* or *fast*: A corpus-based study of English synonyms. *LEARN Journal, 8*(1), 53-62.

Aroonmanakun, W., Tansiri, K., & Nittayanuparp, P. (2009). Thai national corpus: A progress report. *Proceedings of the 7th Workshop on Asian Language Resources (ALR7)*, 153–160. Suntec, Singapore.

Biber, D., Conrad, S., & Reppen, R. (1996). Corpus-based investigation of language use. *Annual Review of Applied Linguistics*, 16, 114-136.

Booij, G. E. (2007). *The grammar of words: An introduction to linguistic morphology*. Oxford University Press.

Corpora: BNC. English. (n.d.). Retrieved January 16, 2022, from https://www.english-corpora.org/bnc/

Hardie, A. (2014, April 28). *Log ratio – An informal introduction*. ESRC Centre for Corpus Approaches to Social Science (CASS). Retrieved January 16, 2022, from http://cass.lancs.ac.uk/log-ratio-an-informal-introduction/

Hong, JF. (2014). Chinese near-synonym study based on the Chinese gigaword corpus and the Chinese learner corpus. In Su, X., He, T. (Eds), *Chinese Lexical Semantics. CLSW 2014*. *Lecture Notes in Computer Science,* vol 8922. Springer, Cham. https://doi.org/10.1007/978-3-319-14331-6_33

Hutto, C. J., & Gilbert, E. (2014). *VADER: A parsimonious rule-based model for sentiment analysis of social media text*. Paper presented at the ICWSM.

Lau, H. Y. P., & Lee, S. Y. M. (2016). A corpus-based study on near-synonymous manner adverbs: "yiran" and "duanran". In Dong, M., Lin, J., Tang, X. (Eds), *Chinese Lexical Semantics. CLSW 2016. Lecture Notes in Computer Science*, vol 10085. Springer, Cham. https://doi.org/10.1007/978-3-319-49508-8_57

Lin, Y.-Y., & Chung, S.-F. (2021). A corpus-based study on two near-synonymous verbs in academic journals: PROPOSE and SUGGEST. *English Teaching & Learning*, 45(2), 189–216.

Mallikamas, P., Chakrabongse, N., & Piammattawat, P. (2010). *Oxford-River books English-Thai dictionary*. Oxford University Press.

Oh, S.-Y. (2000). Actually and in fact in American English: a data-based analysis. *English Language and Linguistics, 4*, 243 - 268.

Pearson Education Limited. (n.d.). *Longman dictionary of contemporary English online*. Retrieved October 9, 2022, from https://www.ldoceonline.com/

Pérez-Paredes, P., & Bueno-Alastuey, M. C. (2019). A corpus-driven analysis of certainty stance adverbs*: Obviously, really* and *actually* in spoken native and learner English. *Journal of Pragmatics*, *140*, 22-32.

Phoocharoensil, S. (2021). Semantic prosody and collocation: A corpus study of the near-synonyms persist and persevere. *Eurasian Journal of Applied Linguistics*, 7(1), 240-258

Royal Institute Dictionary. (2011). Nanmeebooks. (in Thai)

Vongvipanond, P. (1985). *Lexicological significance of semantic doublets in Thai*. The 18th International Conference on Sino-Tibetan Language and Linguistics, August 27-29, 1985, Bangkok, Thailand.

Appendix A

Python Coded for Sentiment Analysis

```python
#!/usr/bin/env python
# -*- coding: utf_8 -*-

## test sentiment of data
import nltk
from nltk.sentiment import SentimentIntensityAnalyzer
nltk.download('vader_lexicon')
import pandas as pd

df = pd.read_csv('input.csv')

df["Score"] = pd.NaT
df["VADER"] = pd.NaT
sia = SentimentIntensityAnalyzer()

for index, row in df.iterrows():
    print(index,'=',row['Word'],row['Context'])
    df.at[index,'Score'] = sia.polarity_scores(row['Context'])["compound"]
    print(df.at[index,'Score'])
    if df.at[index,'Score'] > 0.3:
        df.at[index,'VADER'] = 'POS'
    elif df.at[index,'Score'] < -0.3:
        df.at[index,'VADER'] = 'NEG'
    else:
        df.at[index,'VADER'] = 'NEU'

df.to_csv('out.csv')
```