A CORPUS-BASED STUDY OF GRAMMATICAL COLLOCATION ON ASIAN EFL LEARNERS

Nirma Paris[1], Pratiwi Samad[2], Monsinee Namchan[3]
Universitas Sulawesi Barat[1], Universitas Pohuwato[2], Foreign Relations, Ministry of Public Health, Thailand[3]
nirmaparis@unsulbar.ac.id[1], pratiwisamad@unipo.ac.id[2], monsinee.n@anamai.mail.go.th[3]

Abstract
This article presents a corpus-based study that investigated grammatical collocations. The study aimed to examine and analyze grammatical collocations employed by Asian EFL learners. An Asian learner corpus was built with 38,338 tokens from 20 academic writing papers of Asian EFL learners. The written learner corpus consists of Thai, The Chinese, Indonesian, and The Philippines of academic papers. CLAWS tagger, AntConc (Windows 3.4.4), and manual analysis were used to analyze the data. The results of the study found that grammatical collocations employed by Asian EFL learners were 1,941 tokens of noun + preposition, 467 tokens of verb + preposition, 147 tokens of adjective + preposition, and none of preposition + noun was used in the papers. The study also revealed that the learners used similar prepositions in the top ten (10) frequency which are of, to, in, for, on, as, by, from, with, and about. In addition, the top three (3) frequency of prepositions for each country were of, to, and in. It showed that each country has the same order. However, the order of prepositions for the fourth to the tenth were different.
Keywords: Asian EFL learners, corpus, grammatical collocation, paper.

Introduction

Collocations play a crucial role in learners' L2 fluency and accuracy, language appropriateness and word knowledge. Mehmonova (2022) stated that collocation refers to the potential of two or more words occurring in lexical or syntactic relations. In line with that, Boonraksa and Naisena (2021) pointed out the collocation as the co-occurrence of two or more words to produce new combinations of words with varying meanings depending on the situational context is referred to as collocation. Alsulayyi (2015) claimed that collocation is hard to be mastery by English foreign language

Paris, Nirma. et al. (2023). A Corpus Based Study
of Grammatical Collocation on Asian EFL Learners

Inspiring: English Education Journal
Vol.6 No.2 (September 2023) pp. 151-164
ISSN 2721-2130

learners. They found that it is hard to recognize collocations and understand how words go together (Chi & City, 2023). While Thewissen's (2008) study revealed that EFL learners were difficult to make combinations of preposition. To further, EFL learners often made improper use of prepositions even at advanced stages of their learning (Tahaineh, 2010).

Collocations are important in both language acquisition as well as language production. They assist producers in better expressing ideas and conveying meaning. In truth, employing formulaic language is a question of tradition rather than linguistic requirements. Normally, lifelong exposure to linguistic conventions facilitates conscious acquisition. As a result, first-language learners have no difficulties mastering these important characteristics of language. However, due to their limited exposure to the target language, foreign language learners should be taught these conventional strings of words (Umair et.al., 2023).

There were several previous studies on collocation. For instance, Kuo (2009) investigated errors from 98 writing of EFL learners. The aim of the study was to give suggestions to teachers for helping students work on certain errors. This study focused on the lexical collocation of verb + noun and adjective + noun. The result showed that students at the English intermediate level were struggling with how to use words correctly.

Shokouhi (2010) studied the types of collocation which are lexical and grammatical by using 90-item multiple choice test including noun + noun, noun + verb, verb + noun, and adjective + noun which were lexical collocations type, and noun + preposition and preposition + noun which were grammatical collocation type. The results revealed that lexical collocations are easier to acquire rather than grammatical collocations for learners and from among all subcategories. In addition, noun + preposition was the most difficult while noun + verb was the easiest one.

Alsulayyi (2015) investigated the English essays written by Saudi students majoring in English in the KSA (The kingdom of Saudi Arabia) and those in the UK (The United Kingdom). The study focused on the production of English grammatical collocations amongst these two groups of students. The results showed that Saudi

Paris, Nirma. et al. (2023). A Corpus Based Study
of Grammatical Collocation on Asian EFL Learners

Inspiring: English Education Journal
Vol.6 No.2 (September 2023) pp. 151-164
ISSN 2721-2130

EFL learners in the UK made grammatical collocation errors less than those who learned English in the KSA. In addition, L1 interference played a crucial role in students' erroneous responses, especially those which contained a preposition. For example, the noun + preposition, adjective + preposition and preposition + noun patterns were used incorrectly throughout the essays.

The previous studies on grammatical collocations had some limitation. For instance, (Boonraksa & Naisena, 2021) study did not focus on the verb + preposition collocation (e.g. apply for) and the adjective + preposition collocation (e.g., excited with) which were grammatical collocation type. Besides, several previous studies investigated the two classifications of collocation (grammatical and lexical), yet none of them compare collocation used among Asian leaners (Wu & Tissari, 2021), and (Muhammad Umair, 2023).

According to the statistics of Murti's investigation (Murti, 2023) of corpus-based analysis of grammatical collocations which aimed to find out the comparison between "same with" and "same as", it was found that the term "same as" is far more common than the phrase "same with" due to its higher frequency. Furthermore, the phrase "same with" could be used as a suitable phrase instead of "same as" in at least four contexts or functions. The phrase "same as" used as in relation to, distinctive similarity, as a reference to something, and indicating a state of togetherness. Whereas the phrase "same as" was often used in at least two contexts or functions such as similar circumstances, conditions, situations, or things, and similar acts. Consequently, the present study aims to compare the collocation used among Asian leaners and it focuses on grammatical collocations which are: adjective + preposition, noun + preposition, verb + preposition, and preposition + noun.

The objectives of this study was to examine and analyze grammatical collocation employed by Asian EFL learners. Another objective was to find out any similarities and differences of grammatical collocations employed by Asian EFL learners.

Method

Paris, Nirma. et al. (2023). A Corpus Based Study
of Grammatical Collocation on Asian EFL Learners

Inspiring: English Education Journal
Vol.6 No.2 (September 2023) pp. 151-164
ISSN 2721-2130

This corpus was collected from academic papers of Asian EFL learners studying at Master of Arts in English, Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand. After collecting the academic papers, there were two free software programs used to analyze the data which are CLAWS part-of-speech tagger for English and AntConc. Finally, the data was analyzed manually.

The corpus used in this study called "Asian leaners corpus". It was built from 20 academic papers of 411714 Academic Writing in English Course at Master of Arts in English, Faculty of Humanities and Social Sciences, Khon Kaen University. The papers were legally copyright from the first year of Master of Arts in English students. The writing topic was about the academic writing. The papers consisted of 16 papers from Thai (30,477 tokens), 2 papers from The Chinese (4,038 tokens), 1 paper from Indonesian (1,651 tokens), and 1 paper from The Philippines (2,172 tokens). Therefore, this "Asian learner corpus" consists 38.338 tokens (words) from Thai, The Chinese, Indonesian, and The Philippines of academic papers. In analyzing the data, the references and the figures of each paper were excluded.

After corpus built, CLAWS part-of-speech tagger for English and AntConc were used to analyze the data. The first program was CLAWS part-of-speech tagger for English used to tag the parts of speech. CLAWS (the Constituent Likelihood Automatic Word-tagging System) is the part of speech (POS) tagging or grammatical tagging that is the commonest form of corpus annotation. It is the first form of annotation to be developed by UCREL at Lancaster. The program was the tagging for English text and it has been continuously developed since the early 1980s. The latest version of the tagger, CLAWS4, was used to POS tag approcimately 100 million words of the British National Corpus (BNC).

The second program was AntConc. AntConc is a freeware corpus analysis toolkit for concordancing and text analysis. In the use of AntConc, the present study only focused in the use of keyword list and word list. Keyword list was used to show comparisons between two corpora. While, word list was used to search the most 10 frequency of preposition. Then, this study analyzed the most three frequency of prepositions (of, to, in) on grammatical collocations which are adjective + prepositions

(of, to, in), noun + prepositions (of, to, in), verb + prepositions (of, to, in), and prepositions (of, to, in) + noun.

In analyzing the data, the Corpus of Contemporary American English was used to compare with the results of prepositions of this study. Finally, the data was analyzed manually to investigate words that collocated with prepositions (of, to, in), and check the accuracy of grammatical collocation used with the Cambridge English Dictionaries Online.

Results

The results of this study are presented as following table.

Table 1. The statistic details of the corpus

Asian Corpus	Tokens	Grammatical Collocation	
		Raw	Norm
Thailand	30,477	2,047	67.17
China	4,038	274	67.86
Indonesia	1,651	105	63.60
The Philippines	2,172	129	59.39

The Asian EFL corpus was analyzed in norm of 1,000 tokens because the paper lengths of each country were different. Table 1 provided both raw and norm tokens of grammatical collocations of noun, verb, adjective collocated with *of, to, in* and *of, to, in"* + noun. The results showed that the norm between Thailand (67.17) and China (67.86) were similar, while those countries were different from Indonesia (63.60) and The Philippines (59.39).

There were two freeware programs, *CLAWS tagger* and *Antconc,* used to collect quantitative data. Firstly, the prepositions were sought, and then the data were analyzed by using normalization of 1,000 tokens, words.

Table 2. The top 10 frequency of prepositions

Country			
Thailand	China	Indonesia	The Philippines
of	of	of	of
to	to	to	to
in	in	in	in
for	as	on	as
on	for	for	on
as	with	as	with
by	on	by	for
from	at	with	by
with	from	at	from
about	by	from	about

Table 2 showed the top 10 frequency of prepositions in each country. It found that the first top three frequency of prepositions were the same in each country. The prepositions were "of, to, in", while the rest of the prepositions were also similar, but in the different order.

Table 3. The top three grammatical collocations

Grammatical Collocations	Thailand		China		Indonesia		The Philippines		Total Norm
	Raw	Norm	Raw	Norm	Raw	Norm	Raw	Norm	
Adj. + Prep.	126	4.13	14	3.47	1	0.61	6	2.76	10.97
Verb + Prep.	387	12.70	43	10.65	21	12.72	16	7.37	43.43
Noun + Prep.	1534	50.33	217	53.74	83	50.27	107	49.26	203.61
Prep. + Noun	0	0.00	0	0.00	0	0.00	0	0.00	0.00
Total	2047	67.17	274	67.86	105	63.60	129	59.39	258.01

Table 3 showed that the most frequent grammatical collocations employed Asian EFL learners were nouns + prepositions (203.61), followed by verb +

Paris, Nirma. et al. (2023). A Corpus Based Study
of Grammatical Collocation on Asian EFL Learners

Inspiring: English Education Journal
Vol.6 No.2 (September 2023) pp. 151-164
ISSN 2721-2130

prepositions (43.43), then adjective + prepositions (10.97), and there was none of preposition + noun (0.00).

To know the similarities and differences of grammatical collocations employed by Asian EFL learners, this paper compared the data manually shown in the following table.

Table 4. The similarities and differences of prepositions

No.	Thailand			China			Indonesia			The Philippines		
	Prep.	Token	Norm	Prep.	Token	Norm	Prep.	Token	Norm	Prep.	Token	Norm
1	of	851	32.22	of	123	30.71	of	43	26.04	of	58	26.70
2	to	651	21.36	to	86	20.80	to	34	19.38	to	41	18.88
3	in	545	17.88	in	65	16.10	in	28	16.96	in	30	13.81
4	for	282	9.25	as	31	7.68	on	25	15.14	as	29	13.35
5	on	244	8.01	for	22	5.45	for	11	6.66	on	24	11.05
6	as	244	8.01	with	22	5.45	as	9	5.45	with	12	5.52
7	by	194	6.37	on	20	4.95	by	7	4.24	for	8	3.68
8	from	159	5.22	at	17	4.21	with	5	3.03	by	7	3.22
9	with	130	4.27	from	15	3.71	at	4	2.42	from	6	2.76
10	about	83	2.72	by	12	2.97	from	3	1.82	about	6	2.76
Total		3383	115.5		413	102		169	101.1		221	101.7

The table 4 showed that the first top three frequency of prepositions for each country were in the same order. The first sequence was "preposition of", the second sequence was "preposition to", and the third sequence was "preposition in". The arrangements of prepositions after the third sequence were different, but the prepositions were similar.

Table 5. The most frequent grammatical collocations of adjective + propositions

Collocation	Thailand	China	Indonesia	The Philippines
Adj. + of	hard	N/A	N/A	N/A
	aware	N/A	N/A	N/A

Paris, Nirma. et al. (2023). A Corpus Based Study
of Grammatical Collocation on Asian EFL Learners

Inspiring: English Education Journal
Vol.6 No.2 (September 2023) pp. 151-164
ISSN 2721-2130

	increasing	N/A	N/A	N/A
Adj. + to	able	N/A	N/A	N/A
	important	N/A	N/A	N/A
	similar	N/A	N/A	N/A
Adj. + in	interested	N/A	N/A	N/A
	involved	N/A	N/A	N/A
	successful	N/A	N/A	N/A

The table 5 showed the most frequent grammatical collocations of adjective + preposition. It found that only the papers from Thai learners that had grammatical collocation used of adjective + prepositions occurred more than once. The first was the adjective + prepositions (of) such as *hard, aware,* and *increasing*. For example:

is the ability of the learners to aware of their own learning process, as well
kind of practice helps students to be aware of their learning. It also informs the

The second was the adjective + prepositions (to) such as *able, important, and similar.* For Example:

for as long as humans have been able to reflect on their cognitive experiences, t
the paragraph with the goal of being able to answer the questions she had generated.

The third was the adjective + prepositions (in) such as *interested, involved, and successful*. For example:

interest. Conclusion   Male are interested in economic activities and fact. To con
stly used. Contrastively, female are interested in psychological processes and social c

Table 6. The most frequent grammatical collocation of verb + propositions

Collocation	Thailand	China	Indonesia	The Philippines
Verb + of	consist/base form	N/A	N/A	N/A
	consist/infinitive	N/A	N/A	N/A

Verb + to	need	related	need	N/A
	used	learn	react	N/A
	refers	N/A	contribute	N/A
	tend	N/A	given	N/A
	related	N/A	N/A	N/A
Verb + in	used	N/A	N/A	exposed
	write	N/A	N/A	tended
	use	N/A	N/A	N/A
	applied	N/A	N/A	N/A
	discussed	N/A	N/A	N/A

The table 6 showed the most frequent grammatical collocation of verb + prepositions. It found that there were the same verbs that collocated with prepositions (to) such as *need* and *related*. *Need* found on paper from Thai and Indonesian learners, while *related* found on paper from Thai and The Chinese learners. Example of *need to*:

Thailand

not to use it. For example, students need to recognize that an exam word problem
e sequential processes that learners need to use to aware cognitive activities, and

Indonesia

corrective feedback, the teachers need to choose the suitable feedback for
back to the learners the teachers need to consider any kind of feedback w

Example of *related to:*

Thailand

ng syntactic complexity in writing is related to maturity and new-syntactic-str
elopmental errors are found closely related to each other so that these two

Paris, Nirma. et al. (2023). A Corpus Based Study
of Grammatical Collocation on Asian EFL Learners

Inspiring: English Education Journal
Vol.6 No.2 (September 2023) pp. 151-164
ISSN 2721-2130

China

rate of CET-4 and CET-6 is closely related to the quality of universities, and prc
introductory part is very important, related to the full article, we must seriously

Table 7. The most frequent grammatical collocation of noun + propositions

Collocation	Thailand	China	Indonesia	The Philippines
	use	beginning	aspects	effects
	types	lack	effects	type
Noun + of	effects	part	effectiveness	acquisition
	effect	end	purpose	improvement
	type	form	N/A	degrees
	students	strategies	feedback	feedback
	learners	methods	strategy	attention
Noun + to	writers	students	errors	students
	way	way	students	N/A
	ability	reader	N/A	N/A
	errors	application	feedback	feedback
	students	position	accuracy	difference
Noun + in	self-efficacy	errors	participants	accuracy
	role	rules	N/A	N/A
	approach	skills	N/A	N/A

The table 7 showed the most grammatical collocation of noun + prepositions. It found that there were some of the same nouns that collocated with *of, to* and *in* among learners' papers. The first was noun + preposition (of) such as *effects* and *type.* For example:

Thailand

Paris, Nirma. et al. (2023). A Corpus Based Study
of Grammatical Collocation on Asian EFL Learners

Inspiring: English Education Journal
Vol.6 No.2 (September 2023) pp. 151-164
ISSN 2721-2130

Eslami (2014), investigating the effects of direct and indirect corrective feedback
aning and solving problem. The effects of implicit teaching grammar in second lan

Indonesia

purpose of this is to suggest the effects of written corrective feedback on enhai
vritings. Eslami (2014) studied the effects of direct and indirect written corrective

The Philippines

(2003) studied the differential effects of explicit and implicit corrective feedba
forms in order to measure the effects of the four different tasks mentioned. Fir

The second was noun + prepositions (to) such as *students, way,* and *feedback.* For example:

Thailand

is great if the teachers can help students to find the way to write effectively,
e a helpful tool that encourages students to learn new vocabularies. It is the

China

awareness, teachers can motivate students to watch more English movies and i
s ideas in English writing, to allow students to write the main points on the

Indonesia

le a correction, thereby leaving the students to diagnose and correct it. It can
ffective to the students. They allow students to improve in different ways. Conclu

The Philippines

dback with elicitations that expect students to rethink what they have written .
of written discourse to encourage students to write more. In this study, some

The third was noun + preposition (in) such as *errors, feedback,* and *accuracy.* For example:

Paris, Nirma. et al. (2023). A Corpus Based Study
of Grammatical Collocation on Asian EFL Learners

Inspiring: English Education Journal
Vol.6 No.2 (September 2023) pp. 151-164
ISSN 2721-2130

Indonesia

:ed that many teachers providing feedback in L2 writing context have been concerned
It of the study showed that direct feedback in general helps in EFL learners' better

The Philippines

effects of recasts and metalinguistic feedback in the acquisition of regular past tense
between differing degrees of explicit feedback in accordance to the acquisition of the

Based on the results of this study, it found that collocation of adjective + prepositions, verb + prepositions, and noun + preposition were different from each country of Asian EFL learners. However, there were a few of the same collocated words were used throughout the papers.

Discussion

Compared the result of the study (Asian leaner corpus with the corpus of Contemporary American English), it found that the Asian leaner corpus was similar to Contemporary American English. The top 10 frequency of preposition in Contemporary American English are (of, in, to, for, with, on, at, from, by, and about). While, the Asian learners corpus are (of, to, in, for, on, as, by, from, with, and about). The top three frequency of prepositions of Contemporary American English were "of, in and to", which were similar to the result of this study, but the orders of "in" and "to" were different. However, the proposition "of" was still the most frequency used by both corpora, Contemporary American English and the Asian leaners corpus.

The results of this study was related to Mukundan and Roslim (2009)'s study. There was 153,889 tokens from written corpus. The study aimed to find out the distributions of prepositions and its frequency order. The prepositions' order found was (to, of, in, on, from, at, by, after, before, between, near, under, behind and in front of). The top three frequency of prepositions were "to, of, and in" that is similar to the results of this study in which the top three frequency of prepositions were "of, to, and in".

In relation to the similarities and differences of grammatical collocations employed by Asian EFL learners, it found that collocated words (adjective, verb, noun

Paris, Nirma. et al. (2023). A Corpus Based Study
of Grammatical Collocation on Asian EFL Learners

Inspiring: English Education Journal
Vol.6 No.2 (September 2023) pp. 151-164
ISSN 2721-2130

+ preposition) were different from each country. Nonetheless, if found that there were some of the same collocated words used on the learners' papers. In the collocation of verb + proposition, it found that preposition (to) was similar in using verb "need' + preposition "to". Another example showed in the collocation of noun + preposition, it also found that preposition (to) was also similar in each country. It revealed that papers from each country used the same noun "students" + preposition "to". This finding was along with Mukundan and Roslim (2009)'s study which showed that there were similarities and differences in terms of the co-occurrence of prepositions with other parts of speech.

Conclusion

This study concluded that the grammatical collocations employed by Asian EFL learners at Khon Kaen University were 1,941 tokens of noun + preposition, 467 tokens of verb + preposition, 147 of adjective + preposition, and none of preposition + noun. The study also revealed that the learners used similar prepositions in the top 10 frequency which are of, to, in, for, on, as, by, from, with, and about. In addition, it found that learners from each country applied the same top three frequency of preposition used in grammatical collocation which are *of, to, in*. However, the order of the other prepositions had different order. This study also revealed that the collocated words, (adjective + prepositions, verb + prepositions, and noun + prepositions), were different in each country. Nevertheless, there were a few of the same collocated words used by the Asian EFL learners.

This study was limited to the small size of Asian EFL learners' papers and this study did not investigate the grammatical collocation errors. Further research may investigate the whole types of collocation, both grammatical and lexical collocations. Therefore, it is also suggested to expand the number of learners' corpus in various Asian countries and collect the bigger number of written learners from many countries.

References

Alsulayyi, M. N. (2015). The use of grammatical collocations by advanced Saudi EFL learners in the UK and KSA. *Canadian Center of Science and Education, 5* (1), 32-43.

Paris, Nirma. et al. (2023). A Corpus Based Study
of Grammatical Collocation on Asian EFL Learners

Inspiring: English Education Journal
Vol.6 No.2 (September 2023) pp. 151-164
ISSN 2721-2130

Boonraksa, T., & Naisena, S. (2021). A Study on English Collocation Errors of Thai EFL Students. *English Language Teaching*, *15* (1), 164-177. https://doi.org/10.5539/elt.v15n1p164

Chi, H., & City, M. (2023). Lexical Collocation Errors in Essay Writing: A Study into Vietnamese EFL Students and Their Perceptions. *International Journal of Language Language Instruction, 2* (2), 1–20.

Kuo, C. (2009). An analysis of the use of collocation by intermediate EFL college students in Taiwan. *ARECLS, 6,* 141-155.

Mehmonova, Y. C. (2022). Lexical and grammatical collocations in the english sentences. *Eurasian Research Bulletin 8,* 43-47.

Mukundan, J., Roslim, N. (2009). Textbook representation of prepositions. *English Language Teaching 2* (4), 13-24.

Murti, E. (2023). A corpus-based analysis of grammatical collocations "same as" and "same with*.*"https://repository.unja.ac.id/id/eprint/51664

Shokouhi, H. (2010). Collocation knowledge versus general linguistic knowledge among Iranian EFL learners. *The Electronic Journal for English as a Second Language, 13* (4), 1-10.

Tahaineh, Y.S. (2010). Arab EFL university students' errors in the use of prepositions. *Modern Journal of Applied Linguistics, 2* (1), 76-112.

Thewissen, J. (2008). The phraseological errors of French, German, and Spanish speaking EFL learners: evidence from an error-tagged learner corpus. *Proceedings from The 8th Teaching and Language Corpora Conference (TaLC 8)*, 300-306.

Umair, H. M., Imran, M., Sarwat, S. (2023). An Analysis of the Use of Collocations in English Essay Writing at Undergraduate Level. *Pakistan Languages and Humanities Review*, *7* (1). https://doi.org/10.47205/plhr.2023(7-i)05

Wu, J., & Tissari, H. (2021). Intensifier-Verb Collocations in Academic English by Chinese Learners Compared to Native-Speaker Students. *Chinese Journal of Applied Linguistics*, *44*(4), 470–487. https://doi.org/10.1515/CJAL-2021-0030