

Preposition-Related Collocation use among British and Malaysian Learners: A Corpus Analysis

Ang, L. H.^{1*} and Tan, K. H.²

¹*School of Humanities, Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia*

²*School of Language Studies and Linguistics, Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia*

ABSTRACT

The study examines and compares the use of preposition-related collocations in the writing of Malaysian learners of English and British native speakers of English. The study seeks to answer two research questions: firstly, it sets out to quantify preposition-related collocation use among the Malaysian learners of English and British native speakers of English by measuring the statistical significance of the relevant collocation use in each group using Mutual Information (MI) and a t-score; and secondly, the study aims to identify types of collocational errors associated with prepositions studied in the current research. The frequency-based approach was adopted in the study to define collocations, with the node-and-collocates analysis employed to identify relevant preposition-related collocations. Two references were used to determine the acceptability or otherwise of the collocations: Google Internet search engine results and the online BNCweb corpus. The data revealed that Malaysian learners of English produce more preposition-related collocations than British native speakers of English do. In terms of collocational errors, a stark contrast in the writing of Malaysian learners of English and that of British native speakers of English is apparent, in which preposition-related collocational errors in the Malaysian learner corpus constitute 1% to 7% for certain prepositions, whereas British native speakers' writing was found to be totally free of collocational errors. .

Keywords: Collocational errors, learner corpus, mutual information, node-and-collocates analysis, preposition-related collocations, t-score

ARTICLE INFO

Article history:

Received: 21 September 2015

Accepted: 15 January 2016

E-mail addresses:

leng_hong@hotmail.com (Ang, L. H.),

kimmy@ukm.edu.my (Tan, K. H.)

* Corresponding author

INTRODUCTION

Collocation is becoming increasingly significant in language acquisition as

it has been ascertained that competent use of collocation is a key factor in determining fluent and natural language use (Pawley & Syder, 1983; Sinclair, 1991; Cowie, 1998; Howarth, 1998; Hill, 2000; Nation, 2001; Tan, 2001; Durrant & Schmitt, 2009). In relation to fluency and naturalness of language use, knowledge of collocation also contributes to a learner's communicative competence as collocation is advantageous for processing (Millar, 2011). Recent research has shown that a dearth of appropriate collocations leads to increased and sustained burden on mental processing, which, in turn, could be a barrier to communication (Millar, 2011). This finding is consistent with usage-based models of language acquisition, which grounds language structure in the actual use of language. In view of the importance of collocation in language use and communication *per se*, it is of central importance to understand the notion of collocations, which has also been referred to as formulaic sequences, prefabricated patterns, chunks, clusters, lexical bundles, recurrent sequences and n-grams (Nattinger & DeCarrico, 1992; Stubbs, 1995; Manning & Schütze, 1999; Howarth, 1998; Scott, 2001; Wray, 2002; Schmitt, 2004).

The Notion of Collocation

Studies on collocation have always been conducted within two distinct traditions, which we can refer to as the frequency-based and the phraseology-based traditions. In the former, frequency and statistical measures are integral to explain various phenomena

and instantiations of collocations. In the phraseology-based tradition, collocation is subject to semantic and syntactic analyses, which are apparently less concerned with statistics.

The frequency-based approach to collocation sees collocation as a form of recurrent word combination, which appears more often than by chance. The term collocation was introduced by Firth (1957), who asserted that collocation is crucial in understanding how meaning is created through use at different levels within language. Firth viewed collocations as recurrent sequences of words, where the sequences range from two words up to 15. He insightfully encapsulated the significant role of collocation in language learning by exclaiming, "You shall know a word by the company it keeps" (1957, p. 179). Following Firth (1957), Halliday (1966) developed the notion of collocation by setting the parameters of collocation that limit the co-occurrence of particular words, which facilitates the prediction of word combinations statistically. Halliday has also introduced the terms node, collocate and span, which are still fundamental in frequency-based research at present.

Sinclair (1966; 1991) expanded on Halliday's concept of probability of recurrent word combinations. According to Sinclair, collocation is the occurrence of two or more words within a short span of each other in a text, where a short span is seen as a distance of relevant lexical items (collocates) of the node word. Co-occurrences of node and collocates are usually studied to decide

if they are frequent or not. The notion of collocation, under the frequency-orientated approach, has been expanded by scholars in the field (for example, Stubbs, 1995), who characterised collocation as the occurrence of word combinations greater than by chance in their context and where word pairs are found together more frequently than the occurrence of their component words. The development of the frequency-based approach in collocation research has contributed towards the extension of the notion of collocation, in which collocation is manifested in *lexical bundle* analysis (Biber et al., 1999). Lexical bundles are defined as “recurrent expressions, regardless of their idiomaticity, and regardless of their structural status” (Biber et al., 1999, p.990). The lexical bundle approach allows the study of similar combinations of n-word bundles, for instance 2-word, 3-word, 4-word bundles etc. Corpus-based analyses of lexical bundles commonly focus on texts in specific registers and genres.

The phraseological approach to collocation considers collocation as a form of word combination, which can be delimited from other types of word combinations i.e. free combinations and idioms (Cowie, 1998). A free combination, such as *read the book*, is the least cohesive of all combinations as their components are free to combine with other items. A collocation, such as *commit suicide*, is more restricted in terms of its sense but less frozen than an idiom. An idiom such as *spick and span* is a truly frozen piece of language that has the least complexity.

Frequencies and statistical significance do not play a pivotal role in the phraseological approach as they do in frequency-based tradition. Scholars (for example, Aisenstadt, 1981; Cowie, 1998; Mel’čuk, 1998) viewed collocations as habitually occurring word combinations that are formed by restricted co-occurrence of elements and varying degree of transparency of meaning. The phraseology-based tradition is heavily influenced by research carried out in Russia since the 1940s. It gained popularity in the West from the 1970s onwards, particularly with regards to collocation restriction (Aisenstadt, 1981). Pioneering work within the phraseological approach to collocation include Aisenstadt (1981), Cowie (1998), Howarth (1998), Mel’čuk (1998) and Nesselhauf (2003; 2005).

In view of the phraseology-based approach to collocation, it is worth noting that word combinations differ along a continuum, which makes exact delimitation impossible. This identification of collocation based on a semantically restricted sense invites criticism from the proponents of the frequency-based approach to collocation. Hoey (2005; p. 2) commented that collocations are recurrent combinations that are prevalent in language use and are proven to facilitate the “naturalness” of language production. The motivation of the phraseology-based approach to collocation based on restricted and “semantically anomalous” criteria (Hoey, 2005, p. 16) leaves very frequent and prevalent collocations out of the picture and most probably overlook a number of important

collocations that are indeed prevalent and bear functional value in language use.

Objectives of the Study

The study seeks to use a corpus-linguistic method to examine and compare a Malaysian English learner corpus with a British English learner corpus by focussing on the collocation use of prepositions between these two groups of speakers. The primary aim is to examine the extent of preposition-related collocation use by measuring significant differences of preposition-related collocation use between Malaysian learners of English and British native speakers of English. The second aim is to identify possible types of erroneous preposition-related collocation in both the Malaysian and British native corpora. As prepositions are very frequent in corpora, even in smaller ones, it is deemed suitable to compare and measure statistically the collocation use of prepositions in the two small corpora used in the study. Also, the significant role that preposition plays in formulaic sequences made it a good choice for collocational analysis (Hunston & Francis, 2000). Prepositions have also been notoriously known to pose problems for learners, even for those at advanced level. The approach to the study is from the theoretical and methodological standpoint of corpus linguistics (Sinclair, 1991; Biber et al., 1998). This is to say, the collocations defined in the study are a quantitative and frequency-based phenomenon, which will be observed systematically through the corpus-linguistic analysis of electronically-

stored attested texts. This study seeks to use the corpus-linguistic method to seek answers to the following research questions:

1. Are there any significant differences of preposition-related collocation use between Malaysian learners of English and British native speakers of English?
2. What are the types of erroneous preposition-related collocation found in the writing of Malaysian learners of English and British native speakers of English?

METHODOLOGY

The study focussed on the use of preposition-related collocations defined in the frequency-based tradition represented by the influential pioneer in the field (Sinclair, 1991). In the study, collocation is referred to as word combinations of two or more words occurring near each other in a text. The frequency-orientated approach was adopted in the study as the study sought to measure the extent of preposition-related collocation use as well as to identify possible erroneous preposition-related collocations. The study did not intend to investigate collocations in a semantically restricted sense. It was hoped that such a broader definition of collocation would help to gain deeper insight into the extent of high-frequency collocation use among Malaysian learners of English and British native speakers of English.

Corpora used in the Study

The research compared preposition-related collocations found in the writing of

Malaysian learners of English against those found in native texts. The Malaysian learner corpus used for the study was a sub-corpus of the English of Malaysian School Students (henceforth, EMAS). EMAS is an electronic database compiled by seven researchers from University Putra Malaysia. It contains both written data in the form of essays and oral interviews. The present study only examined the written data in the corpus. The sub-corpus of EMAS used for the study was an untagged learner corpus that contains data in the form of student essays written by 206 students. It consists of 64,692 word tokens and 4,242 word types. The selected data are a compilation of 206 picture-based essays written by Form Four students from three states in Peninsular Malaysia.

The comparative native speaker corpus, Louvain Corpus of Native English Essays (henceforth LOCNESS), was compiled at the University of Louvain la Neuve, Belgium. It comprises essays written by British A-Level and university students as well as by American university students. A sub-corpus of LOCNESS was chosen for the study, which comprises 114 British A-Level student essays, with a total number of 60,398 word tokens and 6,531 word types.

Procedure

Research Question 1: Are there any significant differences of preposition-related collocation use between Malaysian learners of English and British native speakers of English?

The node-and-collocates approach, used to identify the collocations concerned,

was entirely dependent upon computer algorithms to answer Research Question 1. *AntConc* (version 3.2.4w) software was used to perform the relevant frequency calculation and statistical measures. There were three concerns as to how to perform the node-and-collocates analysis. The first concern was how we were to judge if the collocates occurred *significantly frequently* within the span of a given node word. The simplest way to identify frequent collocates is to rank them according to raw frequency figures. Nevertheless, raw frequency is commonly dominated by words from closed grammatical classes such as conjunctions, determiners, prepositions and pronouns, resulting in the difficulty to prove if the collocates and node co-occur significantly and frequently. It is therefore important to have statistical measures that are able to indicate the statistically significant results. Two association measures of collocational strength were performed on each corpus (EMAS and LOCNESS): Mutual Information (henceforth, MI) and the t-score. Essentially, MI highlights the strength of the collocational relationship between the node and collocates, while the t-score normally indicates the degree of certainty that can be claimed about a collocational relationship between the node and collocates in a given corpus (Stubbs, 1995; Barnbrook, 1996). In order to obtain greater statistical significant differences of the collocation use between the Malaysian learners of English and British native speakers of English, a minimum collocate frequency level of 10 occurrences was

applied and the threshold score for both the MI and t-score was set at 4.0. It has been suggested that an MI score of 3 or greater, and a t-score of 2 or more could be held to be significant (Hunston & Francis, 2000; Hunston, 2002).

The second question to ask about node-and-collocates analysis concerned span width, in which the default span of two words to the left and to the right of the node word was set. Restricting the analysis to such a short span of text incurs the danger of missing certain relevant collocations that fall out of the span width. This concern is possibly true. Nevertheless, the present study was concerned with preposition-related collocations, in which the objects (collocates) of the prepositions are always very near to the prepositions (node), as the prepositions are responsible to link the objects to other elements in the sentence environment concerned. Instances of the preposition-related collocations in the corpus used for the study were *shouting for help, some flowers by the riverbank, look*

at the flowers, far away from the girls etc. It is obvious that the preposition-related collocations in the corpus mostly fell within the span of + 2 words of the node word (preposition).

A final question to be answered with regards to the use of the node-and-collocates approach to collocation is that this approach requires a pre-determined list of node words for analysis. The present study focussed on the 10 most frequently used prepositions in both the Malaysian Learner and British native corpora. As can be seen in Table 1, the same top 10 prepositions were comparable in both corpora, with the omission of *into* from EMAS and *but* from LOCNESS. The omissions were deemed necessary as *into* does not fall into the top 10 preposition ranking in LOCNESS, while *but* is not included in the top 10 prepositions in EMAS. It should also be noted that collocations which were found erroneous were omitted in the list of collocations as they were not valid quantitative data.

Table. 1
Rank and Frequency Data for the Top 10 Prepositions in EMAS and LOCNESS

EMAS			LOCNESS		
Rank	Word	Frequency	Rank	Word	Frequency
2	to	2494 (+25.32%)	2	to	1990
8	of	857	3	of	1856 (+116%)
10	for	688 (+11%)	7	in	1054 (+104%)
16	in	516	11	for	620
24	at	511 (+190%)	13	as	564 (+57%)
36	as	359	20	on	378 (+32%)
42	with	263	23	with	307 (+17%)
49	on	286	25	by	306 (+50%)
57	from	216	35	from	218 (+1%)
59	by	204	46	at	176

Research Question 2: What are the types of erroneous preposition-related collocation found in the writing of Malaysian learners of English and British native speakers of English?

To answer Research Question 2, two references were used to determine the acceptability of the collocations. The preposition-related collocations identified through the node-and-collocates analysis were checked against the GoogleTM Internet search engine and the online BNCweb to validate their acceptability or otherwise. Collocations were judged acceptable if they were found in identical form in both the GoogleTM Internet search results and the online BNCweb corpus.

FINDINGS AND DISCUSSION

Research Question 1: Are there any significant differences of preposition-related collocation use between Malaysian learners

of English and British native speakers of English?

The results of the MI analysis are presented in Table 2. This analysis found that there were more preposition-related collocation types in EMAS than there were in LOCNESS in five out of 10 prepositions studied (*for, at, with, from* and *by*). Only three prepositions (*to, of* and *in*) reversed this trend. The remaining two prepositions (*as* and *on*) bore the same number of collocation types in both EMAS and LOCNESS. In terms of the collocate tokens, the majority of prepositions (*to, of, for, at, with, on, from* and *by*) in EMAS outnumber those (*in* and *as*) in LOCNESS.

The results of the t-score analysis were similar to that of the MI, as shown in Table 3. The slight difference is that the preposition-related collocation types in EMAS (*to, for, at, as, with, on, from* and *by*) substantially outnumber those in LOCNESS (*of, in*) in the ratio of 8 to 2.

Table. 2
Collocation Types and Tokens Identified by Mutual Information Analysis

	EMAS		LOCNESS	
	Types	Tokens	Types s	Tokens
to	29	1549 (+134%)	33 (+14%)	661
of	14	676 (+60%)	23 (+64%)	423
for	15 (+67%)	531 (+252%)	9	151
in	7	137	20 (+186%)	323 (+136%)
at	14 (+600%)	355 (+1379%)	2	24
as	8	519	8	865 (+67%)
with	3 (+200%)	49 (+345%)	1	11
on	5	111 (+50%)	5	74
from	6 (+100%)	100 (+186%)	3	35
by	2 (+100%)	59 (+354%)	1	13

Table. 3
Collocation Types and Tokens Identified by t-Score Analysis

	EMAS		LOCNESS	
	Types	Tokens	Types s	Tokens
to	74 (+64%)	5642 (+100%)	45	2827
of	23	1632	34 (+48%)	2579 (+58%)
for	18 (+125%)	1109 (+144%)	8	454
in	7	570	22 (+214%)	1188 (+108%)
at	12 (+1100%)	701 (+874%)	1	72
as	13 (+18%)	753	11	1194 (+59%)
with	6 (+200%)	196 (+48%)	2	132
on	6 (+50%)	277 (+5%)	4	265
from	4 (+300%)	156 (+71%)	1	91
by	4 (+100%)	198 (+41%)	2	140

In summary, the node-and-collocates analysis above illustrates that Malaysian learners of English significantly use more preposition-related collocations than do British native speakers of English. In dealing with preposition-related collocations, it should always be borne in mind that prepositions are abundant in language and preposition-related collocations are highly frequent and common in every English variety as prepositions are important function words in the English Language. The results obtained in the present study, though statistically significant, only prove that Malaysian learners of English (non-native speakers) use more preposition-related collocations, which is very frequent in nature, in their writing when compared to native speakers of English. The results of the study do not explicitly or implicitly offer evidence of overuse of preposition-related collocations by Malaysian learners of English as the study was aimed at measuring the whole picture of preposition-related

collocation use among Malaysian learners of English and British native speakers of English, and it did not intend to partition the collocation types and tokens into groups of overuse and underuse.

Research Question 2: What are the types of erroneous preposition-related collocation found in the writing of Malaysian learners of English and British native speakers of English?

It was deemed necessary to establish how many preposition-related collocations are qualitatively unacceptable, which would be excluded as valid quantitative data. Also, learner language has been found to be collocationally error-prone (Nesselhauf, 2005; Siyanova & Schmitt, 2008; Millar, 2011), which made it necessary to identify the possible errors associated with the types of collocation concerned.

The results of identification of collocational errors for the 10 prepositions in EMAS and LOCNESS are shown in

Table 4. A stark contrast between EMAS and LOCNESS is clearly apparent, in which preposition-related collocational errors in EMAS constituted 1% to 7% in each type of preposition, whereas LOCNESS was found to be totally free of collocational errors. Collocational errors in EMAS were found associated mostly with the prepositions *for*, followed by *to*, *at* and *on*. To gain a deeper insight into the types of collocational errors concerning prepositions, a further classification of the preposition-related collocational errors is shown in Table 5.

Altogether, 130 instances were found in EMAS, of which most involved the superfluous use of prepositions (95% of preposition errors), followed by wrong choice of preposition (5% of preposition

errors). The superfluous prepositions in the collocations concerned seemed to be quite systematic as they involved mainly two prepositions: *to* and *for*. The systematised errors could be the evidence of fossilisation in the writing of the Malaysian learner of English. It is worth noting that the prevalence of superfluous prepositions in the collocations indicates that Malaysian learners of English overuse certain prepositions and, at the same time, they are uncertain about the correct use of these prepositions. The finding of this research question is consistent with the results of some previous research (for example, Ang et al., 2011). It might be academically worthwhile to investigate the patterns of erroneous preposition-related collocations

Table. 4
Percentage Frequencies of Collocational Errors for 10 Prepositions in EMAS and LOCNESS

	Examples	EMAS	LOCNESS
to	They love to fishing at the river.	3%	0%
of	-----	0%	0%
for	Let's go for fishing.	7%	0%
in	-----	0%	0%
at	The scene at there was so beautiful.	2%	0%
as	-----	0%	0%
with	-----	0%	0%
on	Ramu invited me to fishing on the river.	1%	0%
from	-----	0%	0%
by	-----	0%	0%

Table. 5
Types of Collocational Errors in Preposition-Related Collocations

Types of preposition error	Occurrences (tokens)
Wrong choice of preposition	6
Superfluous preposition	124
Total	130

in future research to determine if cross-linguistic influence takes place here.

To recapitulate, EMAS was found to contain four types of preposition-related collocational error (*for*, *to*, *at* and *on*). On the other hand, LOCNESS, the British native speakers' corpus, comprised collocationally well-formed preposition-related collocations, rendering it error-free collocationally. The findings of the study present the fact that learner language, at least in the Malaysian L2 context, is riddled with errors, which are overt even in small corpora such as the one used for the current study.

CONCLUSION

The study employed tools and methods of corpus linguistics to examine prepositional-related collocation use in the writing of Malaysian learners of English and British native speakers of English. Two research questions were set and answered through the node-and-collocates analysis as well as frequency and statistical counts. Firstly, Malaysian learners of English significantly used more preposition-related collocations than British native speakers of English did; and secondly, preposition-related collocational errors were prevalent and overt in the writing of Malaysian learners of English, associated particularly with prepositions such as *for*, *to*, *at* and *on*, while the writing of British native speakers of English was error-free collocationally.

It should be borne in mind that these findings are based on the analysis of small corpora: a sub-corpus of EMAS and a sub-

corpus of LOCNESS. It should also be pointed out that the collocations studied are taken from one word class, the preposition. Generalisations made in the study are subject to confirmation or challenge by future research that may look at larger corpora and examine different sets of collocations.

Lastly, learners should be encouraged to improve their knowledge and use of collocations as the importance of collocations in determining fluency, naturalness and effective communication has been established by research conducted in the field. It is therefore vital for educators to expose learners to real language in use, which is advocated by usage-based models of language. Language teachers may exploit the available resources, such as Internet resources and linguistic tools to guide learners to learn and develop their knowledge of collocations in real-life situations.

REFERENCES

- Aisenstadt, E. (1981). Restricted collocations in English lexicology and lexicography. *Review of Applied Linguistics*, 53, 53-61.
- Ang, L. H., Rahim, H. A., Hua, T. K., & Salehuddin, K. (2011). Collocations in Malaysian English learners' writing: A corpus-based error analysis. *3L: The Southeast Asian Journal of English Language Studies*, 17(Special Issue), 31-44.
- Anthony, L. (2015). *Antconc (3.4.1w)*. Japan: Waseda University.
- Barnbrook, G. (1996). *Language and computers*. Edinburgh: Edinburgh University Press.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). Lexical expressions in speech and writing. In *Longman grammar of spoken and written English* (pp. 988-1036). Harlow, Essex: Longman.
- BNCweb online corpus. (2015, August 25). *BNCweb online corpus*. Retrieved from <http://bncweb.info/>
- Cowie, A. P. (1998). *Phraseology: Theory, analysis, and applications*. Oxford: Oxford University Press.
- Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *IRAL-International Review of Applied Linguistics in Language Teaching*, 47(2), 157-177.
- Firth, J. R. (1957). *Papers in linguistics 1934-1951*. London: Oxford University Press.
- Google Internet search engine. (2015, August 25). *Google Internet search engine*. Retrieved from <http://www.google.com>
- Halliday, M. A. K. (1966). Lexis as linguistic level. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, & R. H. Robins (Eds.), *In memory of J. R. Firth* (pp. 148-162). London: Longman.
- Hill, J. (2000). Revisiting priorities: From grammatical failure to collocational success. In M. Lewis (Ed.), *Teaching collocations: Further development in the lexical approach* (pp. 47-69). London: Language Teaching Publications.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge.
- Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 161-186). Oxford: Oxford University Press.
- Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins Publishing Company.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Manning, C., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.
- Mel'čuk, I. (1998). Collocations and lexical functions. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 161-186). Oxford: Oxford University Press.
- Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics*, 32(2), 129-148.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223-242.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins Publishing Company.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards, & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-226). London: Longman.
- Schmitt, N. (2004). *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins.
- Scott, M. (2001). Comparing corpora and identifying key words, collocations and frequency distributions through the Wordsmith Tool Suite of computer software. In M. Ghadessy, A. Henry, & R. Roseberry (Eds.), *Small corpus studies and ELT* (pp. 47-67). Amsterdam: John Benjamins.
- Sinclair, J. (1966). Beginning the study of lexis. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, & R. H. Robins (Eds.), *In memory of J. R. Firth* (pp. 410-430). London: Longman.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review*, 64(3), 429-458.
- Stubbs, M. (1995). Collocations and semantic profiles. On the cause of the trouble with quantitative studies. *Functions of Language*, 2(1), 23-55.

Ang, L. H. and Tan, K. H.

Tan, K. H. (2001). *Corpora: Characteristics and related studies*. Malaysia: UKM Press.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.