Jurnal Educatio FKIP UNMA
Volume 6, No. 2, December 2020, pp. 703-710
DOI:https://doi.org/10.31949/educatio.v6i2.757

P-ISSN 2459-9522
E-ISSN 2548-6756

LEXICAL COLLOCATION PRODUCTIVITY OF INDONESIAN L2 WRITERS IN ESSAY: A COMPARATIVE CORPUS-BASED STUDY

Yenni Arif Rahman
Universitas Bina Sarana Informatika
yeni.yar@bsi.ac.id

ABSTRACT

The richness of collocation usage reflects the language mastery of English users. However, it has been recognized that L2 users often have problems with collocations due to several reasons. This study reports on lexical collocation productivity of Indonesian L2 writers to English-native writers in essays. The corpora were taken from 20 essays written by Indonesian L2 writers and English-native writers in English newspaper opinion column. To conduct the analysis, this study employed corpus-based comparative analysis suggested by Gonzales and Ramos. This is done by extracting all lexical collocation from the text by utilizing AntConc, a corpus analysis software. Then, collocations were sorted out from free combinations and collocation errors by using https://skell.sketchengine.eu, a reference corpora search-engine. The average use of lexical collocation of Indonesian L2 writers in essays was compared with lexical collocation of English-native writers. The results showed that Indonesian L2 writers is less productive than English-native writers in utilizing lexical collocation in their essays. Of the 4481 token in Indonesian L2 essays, there were 226 collocation in use or 50 collocations per 1000 token. That result was much lower than English-native collocation in essays which reports 80 collocations per 1000 token or 320 collocations of 3968 token.

***Keywords**: Lexical Collocation Productivity, Indonesian L2 Writers, Corpus-Based Study*

ABSTRAK

Kekayaan penggunaan kolokasi mencerminkan penguasaan pengguna bahasa Inggris. Namun, diakui bahwa pengguna Bahasa Inggris sebagai bahasa kedua (L2) sering mengalami masalah dengan kolokasi karena beberapa alasan. Penelitian ini melaporkan produktivitas kolokasi leksikal penulis L2 Indonesia dibanding penulis Inggris sebagai penutur asli (L1) dalam esai. Korpora diambil dari 20 esai yang ditulis oleh penulis L2 Indonesia dan penutur asli (L1) Inggris di kolom opini surat kabar berbahasa Inggris. Dalam proses analisisnya, penelitian ini menggunakan analisis komparatif berbasis korpus seperti yang disarankan oleh Gonzales dan Ramos. Prosedur analisis dilakukan dengan mengekstraksi semua kolokasi leksikal dari teks dengan menggunakan AntConc, perangkat lunak analisis korpus. Kemudian, kolokasi dipisahkan dari frasa bebas dan kesalahan kolokasi dengan menggunakan https://skell.sketchengine.eu, sebuah mesin pencari rujukan korpora. Kemudian rata-rata penggunaan kolokasi leksikal penulis L2 Indonesia dalam esai dibandingkan dengan kolokasi leksikal penutur asli (L1) Inggris. Hasil penelitian menunjukkan bahwa penulis L2 Indonesia kurang produktif dibandingkan penutur asli (L1) Inggris dalam penggunaan kolokasi dalam esai. Dari 4481 token pada esai L2 Indonesia, terdapat 226 kolokasi yang digunakan atau 50 kolokasi per 1000 token. Hasil itu jauh lebih rendah daripada kolokasi penutur asli (L1) Inggris dalam esai yang menggunakan 80 kolokasi per 1000 token atau 320 kolokasi dari 3968 token.

Kata Kunci: Produktivitas Kolokasi Leksikal, Penulis L2 Indonesia, Studi Berbasis Korpus

Submitted Nov 25, 2020| Revised Des 28, 2020| Accepted Des 31, 2020

Introduction

The use of appropriate collocation is a reflection of the naturality of language. The thesis has been justified in daily speech acts where collocation is one of the lexicons used frequently as dictions by native (Nation, 2001) In other word, the use of productive and accurate collocation determines the level of a foreign language learner. However, previous studies have found that the foreign language learners pose difficulty to use appropriate collocations due to the word combinations which don't provide particular pattern to indicate whether they are collocated (Hashemi et al., 2012). This feature drives collocation as one of the most challenging field to learn for foreign language learners.

The empirical study conducted by Howarth (1998) shows that the main cause of the difficulty of learners using collocation is based on collocation knowledge. The difficulty originated from English as formulaic language is generally agreed to be acquired through exposure. This mean English is language that consist of chunk of phrases that make up the sentense. Eventhough certain factors such as materials, teachers, and learners also play important role in the process (Webb & Kagimoto, 2012). However, literature has shown that there are some challenges which may inhibit the learners from proficiently acquiring formulaic language simply from the input. Further, Laufer and Waldman (2011)reported in his research that non-native English Learners don't use collocation as much native speakers do. The study conducted by Paquot and Granger (2012) also reinforces the finding that the non-native language learners likely underuse "repetitive word combinations that were most academically similar". Nation (2001) also added that collocations also contain some degree of semantic unpredictability which caused learners who lack of collocative knowledge will consequently fail to produce proper word usages. He further argued that collocations reflect the fluency of language learners and as a mark of language mastery.

From the premises, this study aims to explore the extent of Indonesian L2 writers in using lexical collocations in academic writing setting. The study also employs native speaker of English collocations productivity in writing as a standard measurement. The result determines whether the lack of collocational usage occurs in indonesian L2 context as reported by Laufer & Waldman and Paquot & Granger in other L2 samples. If does, then it is relatively accepted that the tendency of collocation as one of challenging vocabulary development even for advanced English learners may pique the different approach to the learning attitude and pedagogy.

As the basis of the collocation and phraseology study, Firth (1957) first coined the term 'collocation' as word combinations that co-occur frequently. Others like Lewis (1993) and Nation (2001) define collocation as combination of words that appear naturally greater than random frequency. From those definitions, we can substract that collocation have charactertics to appear frequently in combination so it is conventionally accepted by native speaker.

Jafarpour (2013) pointed out that most recent research has endeavored to explore grammatical collocations, while much less is done on lexical collocations. However, Bahns and Eldaw (1993) indicated that lexical collocations are especially problematic for L2 and foreign learners. This is mainly caused of lexical combination flexibility whose the combination can be replaced with other similar words. This loose combinations sometimes drive the experienced L2 writers feel insecure to combined predictable collocated words. The insecurity caused most of the writers prefer to play safe with their word choices. Thus the combinations is not as rich as native speakers.

In terms of strength, English collocations can be classified into three: strong, fixed, and weak collocations (Shammas, 2013). Strong collocations refers to the words which are very closely associated with each other. Thus it rarely collocates with any other word. the word like *mitigating* always collocates with *circumstances* and *factor*. Fixed collocations are collocations so strong that they cannot be changed in any way. For example, you can say *I was walking to and fro*. No other words can replace *to* or *fro* or *and* in this collocation. It is completely fixed. The meaning of some fixed collocations cannot be guessed from the individual words. These collocations are called idioms. Weak collocations are made up of words that collocate with a wide range of other words. For example, you can say you are *in broad agreement* with someone. However, *broad* can also be used with a number of other words – *a broad avenue, a broad smile, broad shoulders, a broad accent, a broad hint* and so on. These are weak collocations, in the sense that *broad* collocates with *a broad range* of different nouns. Strong collocations and weak collocations form a continuum, with stronger ones at one end and weaker ones at the other. Most collocations lie somewhere between the two. For example, the adjective *picturesque* collocates with *village*, *location* and *town*, and so appears near the middle of the continuum.

In terms of its contruction, Benson et al. (1986) classified collocations into lexical and grammatical. In this study, the discussion solely focuses on lexical collocations which are composed of content words. In lexical cohesion, the five main parts of speech, verb, noun, adjective, adverb, and prepositions forms a predictable connection one after another. Thus the possible combinations of lexical collocations are mention in table 1.

Table 1. Lexical Collocation Combinations

Type	Pattern	Example
L1	Verb + Noun Phrase/Pronoun/Prepositional Phrase	set a record (verb + noun phrase)
L2	Verb + Noun	commit suicide
L3	Noun + Verb	bomb explode, lions roar
L4	Verb + Adverb	apoligize humbly
L5	Noun + Noun	a piece of advice
L6	Adverb + Adjective	completely satisfied
L7	Adjective + Noun	strong tea, excruciating pain

(adapted from Benson et al. (1986))

Based on the list, the comparative study identifies those lexical collocations in each essay sample. Then the result will be categorized according to the type of combination which are L1 until L7.

Method

The Target Corpus

The target corpus of this study are collocations exhibit in essays written by both Indonesian and English-native (L1) writers in English-written newspapers. The samples consist of 10 essays in opinion column with two English newspapers: Jakarta Post and NewYork Times. Both newspapers represent Indonesian L2 essays and English-native (L1) essays respectively with various types discussion range from politics, education, and economic issue. The comparative analysis of the target corpus will be head-to-head according to genre to avoid bias in analysis.

The token of the collected text will be calculated automatical by using AntConc Software. While the word type is ignored and its apperance in token doesn't influence the number of collocation appear in token. The incorrect collocations in the corpus will be ignored since it doesn't fulfil the standard and necessity of the study.

The Reference Corpora

Both category of target corpus of this study, Indonesian L2 essays and English-native (L1) essays, were compared with https://skell.sketchengine.eu/#home?lang=en. This web engine specializes in collecting English phrases and collocations of several major English versions like American English and British English. https://skell.sketchengine.eu/#home?lang=en backups their list with several major corpora as database (https://www.sketchengine.eu/corpora-and-languages/corpus-list/) like: Corpus of Contemporary American English (COCA list), The Academic Word List/AWL, and the British Academic Written English (BAWE). Its complete database ensures the precise collocation, either weak, strong or fixed collocation, with the number of word frequency appears in each searching query. Once the corpora are listed then one by one they are put in the searching engine to match the collocation profile.

Figure 1. Skell Reference Corpora

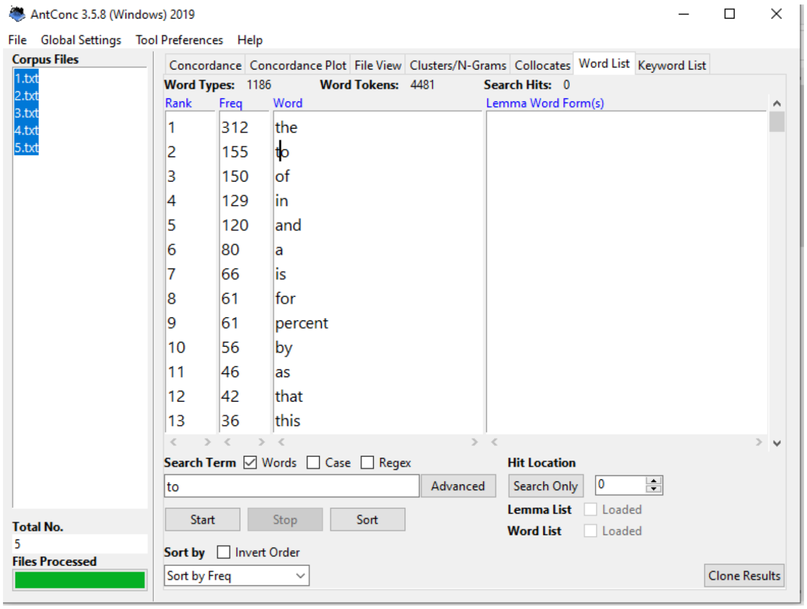
The Comparative Procedure

The study adapeted comparative corpus-based procedure suggested by González and Ramos (2016). The comparative procedure of data collection involved several steps which include: selecting the essay in English written newspaper, converting data if necessary into text before they were ready for analysis. The next step involved corpus cleaning in which the converted texts as raw data were "cleaned" from typos and unnecessary information such as quote and references if any.

The most important step of this study is cleaning unneccasary phrases which doesn't include in table 1. collocation combination. The procedure includes inserting the clean text in the AntConc as a software corpus analysis toolkit. The software counts the token and word type and then analyze the text according to the neccesity of the study. In the wordlist tab of this software, it shows the frequency of word listed in the text and it can be used to predict the phrases that belong to lexical collocation combination. for example if the word "learning" appear in the tab list. The researcher then clicks the word in the tab and the software will show the cooperated word in the text. This is done manually. The text processing is shown in figure 1.

The corpora were then collected in the matrix and then inputted in the reference corpora software web-engine, Skell.skecthengine.eu, to check their collocability. The phrases which are not collocated or contain error are ignored and expelled from the list. The corpora of each essay category are recapitulated to find the general trend of the analysis. This procedure is applied to both essay category: Indonesian L2 essays and English -native (L1) writer essays. To find the general trend, both cetegory are compared to show the productivity of each category.

Figure 2. Corpus Text Processing of Indonesian L2 Essays



AntConc 3.5.8 (Windows) 2019

Word Types: 1186 Word Tokens: 4481 Search Hits: 0

Rank	Freq	Word
1	312	the
2	155	to
3	150	of
4	129	in
5	120	and
6	80	a
7	66	is
8	61	for
9	61	percent
10	56	by
11	46	as
12	42	that
13	36	this

Search Term ☑ Words □ Case □ Regex

to

Hit Location: Search Only 0

Lemma List □ Loaded

Word List □ Loaded

Total No.
5

Files Processed

Sort by □ Invert Order

Sort by Freq

The Collocation Productivity Calculation

The collocation productivity of both selected categories, Indonesia L2 writer essays and English-Native writer essays, are measured by using the mean factor (Granger & Bestgen, 2014). The mean factor is interpreted as the average use of collocations in token (the overall sum of word in texts, L1-L7 per 1000 tokens). It works with the summation of all lexical collocation found in texts, then divided by how many tokens in the texts. This calculation runs to two categories, Indonesian L2 collocations and English-native collocations. The formula is presented as follow:

$$\text{Collocation Productivity} = \frac{\sum \text{L1-L7 x 1000}}{\text{Token}}$$

Remark:
∑L1-L7: the sum of L1 to L7
Token: the total number of word presents in the text

Finding and Discussion

To answer the research question of collocation productvity of Indonesian L2 writers in essays, the research focused on comparing the result of Indonesian L2 collocation productivity with English-native collocation productivity. The first step of the procedure is to find the number of token of both selected category. The 5 clean texts of each selected category are inserted in AntConc software tool kit to count the token and word type. The token is used as one of the free variable altogether with lexical collocation sum to find the average of productivity (bound variable). The word type can be ignored in this regard.

Of the 5 texts in the Indonesian L2 essays, the number of token scores 4481 token and 1186 word type. This is the accumulation of five texts (see figure 2). The number of token and word types from text 1 until text 5 is presented respectively: 851 token with 351 word types, 879 token with 355 word types, 1005 token with 337 word types, 1112 token with 419 word types, and 634 token with 277 words types. Of the 5 texts in English-native essays, the number of token scores 3968 token with 1377 word types. The number of token and word types from text 6 until text 10 is presented respectively: 641 token with 332 word types, 762 token with 377 word types, 797 token with 401 word types, 844 token with 417 word types, and 924 token with 425 word types.

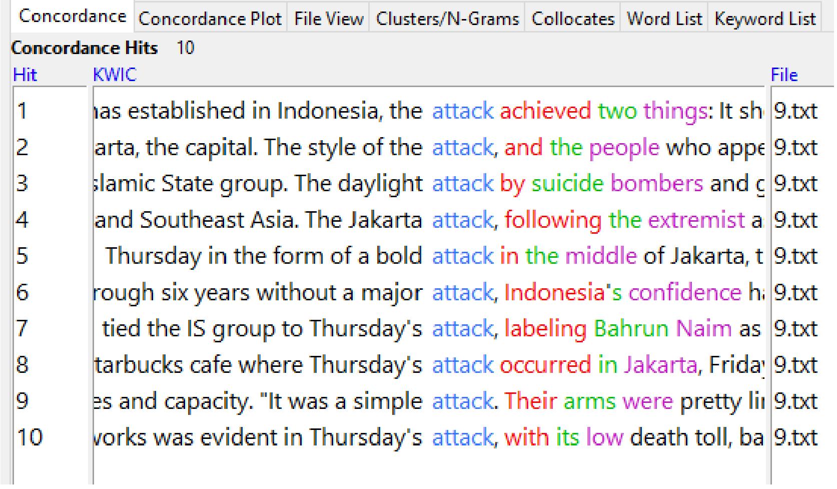
Following the procedure of the research, the next step is finding the number of collocation of both categories, Indonesian L2 essays and English-native essays. The study has found 226 collocation after sorting out 253 phrases of text 1 to text 5 with skell.sketchengine.eu/#home?lang=en. Some of collocation sampels are presented in table 2. In collocation samples of text 1, the word 'online' has 15 concordance hit. It means it collocates with some 15 lexis in the text. The collocations like, online learning, online classes, teach online, online teaching, online universities, and online programs dominates the text. In collocation samples of text 2. The word 'financial'has 14 concordance hit. It means it collocates with 14 lexis in the text. The collocations like, financial access, financial product, financial sector, non-financial institutions, financial inclusion, and financial terms, dominate the text. Other collocations are presented in table 2.

Table 2. Collocation sample of Indonesian L2 Essays

Collocation samples of Text 1	Collocation samples of Text 2	Collocation samples of Text 3	Collocation samples of Text 4	Collocation samples of Text 5
online learning	economic growth	price boom	painful impact	huge demand
video call	development goal	tax revenues	remain sluggish	meet the need
big cities	rapid growth	economic growth	global recovery	poverty line
best alternative	poverty rate	economic-downturn	production costs	housing needs
keeping schools	serious level		unduly burdened	loan liquidity
learning material	financial access	sluggish trend	labor costs	housing loan
regular classroom	low income	healthy margin	fairly cheap	low interest
common reason	financial sector	commodity prices	raw materials	housing supply
having friend	banking products	grows abundantly	expensive cost	state budget
stuck alone	transfer facilities	meet demand	textile industry	legal basis
general consensus	financial access	national average	Take place	custodian bank
highly beneficial	operational costs	limited number	provisional	foreign worker
good alternative	remote areas	highly beneficial	free trade	labor cost
novel concept	micro credit	mining sectors	investment plan	additional cost
silver lining	lending portfolio		non-tariff policies	drive the demand

In the part of English-native essays, the study has found 320 collocation after sorting out 372 phrases of text 1 to text 5. Some of collocation sampels are presented in table 3. In collocation samples of text 6, the word 'impeachment' has 11 concordance hit. However not all hits are collocation, some of them just make the common phrase which exclude collocation. This has been tested in https://skell.sketchengine.eu/#home?lang=en. The word 'impeachment' collocate 'voter' as collocation. In collocation samples of text 7. The other collocations like real candidate, bridge divide, barely functioning, etc appear once or twice in the text . In collocation samples of text 8, the word 'shooting' collocates with words like mass and rampage. It appears 6 times in the text. Other collocation of text 8 like toougher action, against terrorism, or take credit, appear once or twice in the text.

Figure 3. Collocation Analysis of Text 9



Concordance Hits 10

Hit	KWIC	File
1	ıas established in Indonesia, the attack achieved two things: It sh	9.txt
2	arta, the capital. The style of the attack, and the people who appe	9.txt
3	;lamic State group. The daylight attack by suicide bombers and ɡ	9.txt
4	and Southeast Asia. The Jakarta attack, following the extremist a	9.txt
5	Thursday in the form of a bold attack in the middle of Jakarta, t	9.txt
6	rough six years without a major attack, Indonesia's confidence h	9.txt
7	tied the IS group to Thursday's attack, labeling Bahrun Naim as	9.txt
8	tarbucks cafe where Thursday's attack occurred in Jakarta, Frida	9.txt
9	es and capacity. "It was a simple attack. Their arms were pretty li	9.txt
10	vorks was evident in Thursday's attack, with its low death toll, ba	9.txt

In collocation samples of text 9 (see figure 3), the word 'attack' create collocation such as: major attack, attack occur, bold attack, daylight attack, and inspire attack. Other collocations like islamic militants, suicide bombers, emerging threat etc. appear once or twice in the text 9 (see table 3). In collocation samples of text 10, the word 'election' collocates with word 'stolen'. The other collocations like violent attack, electoral votes, and security forces appear once or twice in the text.

Table 3. Collocation sample of English-native Essays

Collocation samples of Text 6	Collocation samples of Text 7	Collocation samples of Text 8	Collocation samples of Text 9	Collocation samples of Text 10
breathing space	rare candidate	mass shooting	stand near	violent attack
political price	bridge divide	tougher action	attack occurred	seditious rhetoric
legal pathway	barely functioning	against terrorism	major attack	electoral votes
widely criticized	political climate	take credit	Islamic militants	security forces
stalling ploy	biggest audience	choose words	bold attack	raise a finger
conservative-	found a successor	gun violence	daylight attack	permission slip
party	carry forward	began weighing	suicide bomber	similar opposition
opposition-	adoring audience	open fire	gunmen target	beyond the pale
parties	gushed over	critically wound	emerging threat	bear a measure
court reviews	former rival	began trickling	scant details	vigorous retailing
political situation	possible successor	get tough	beyond saying	public confidence
ruling party	mere fact	deadly attacks	domestic militant	conclude violence
legal roadmap	stark reminder	shooting rampage	launch attacks	security barriers
initial support	recession doldrum	temporary ban	scare tactics	presidential
state prosecutors	health overhaul	political	developing nation	election

huge protests	economic reform	prospects	moderate Islam	reasonable doubts
ruling party	gun restrictions	appear jarring	took a battering	factual basis
conflicting views	animate voters	deep frustration	sustained effort	decried efforts
disparate groups		terror threats	death toll	sow the wind
		horrific act		swore oaths

The final step is counting the productivity of both categories with the formula provided in the method. The process of counting the variables is presented in table 4.

Table 4. The collocation Productivity of Indonesian L2 Writers in Essay

The free variables	figure	Result
∑L1-L7:	226	Collocation Productivity $= \frac{\sum L1\text{-}L7 \times 1000}{Token}$ $= \frac{226 \text{ X } 1000}{4481}$ $= 50,5$
Number of Token:	4481	

Table 5. The Collocation Productivity of English-Native Writer in Essay

The free variables	figure	Result
∑L1-L7:	320	Collocation Productivity $= \frac{\sum L1\text{-}L7 \times 1000}{Token}$ $= \frac{320 \text{ X } 1000}{3968}$ $= 80,6$
Number of Token:	3968	

From the result of table 4, it can be interpreted that the collocation productivity of Indonesian L2 writer in essay is around 50 collocation per 1000 token. While From table 5, it can be concluded that the collocation productivity of English-Native Writer in Essay is 80 collocations per 1000 token.

Conclusion

Based upon the finding of table 4 and table 5, it can be inferred that Indonesian L2 writers is less productive than English-Native Writer in utilizing the collocation in their essays and the proportion of productivity is much lower than collocation produced by English -native writers by almost a half. In other words, L2 English writers of Indonesian still find difficulties in the production of collocations. This finding confirms what Howart (1998) said that English as a formulaic language restricts L2 learners to acquaire vocabularies as many as natives do.

References

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, *21*(1), 101–114.

Benson et al. (1986). *Lexicographical description of English*.

Firth, J. R. (1957). *A synopsis of Linguistic Theory, 1930-1955. Studies in Linguistic Analysis* (pp. 1–31). Blackwell.

González, A. O., & Ramos, M. A. (2016). A Comparative Study of Collocations in a Native Corpus and a Learner Corpus of Spanish A Comparative Study of Collocations in a Native Corpus and a Learner Corpus of Spanish. *Procedia - Social and Behavioral Sciences*, *95*(October 2013), 563–570. https://doi.org/10.1016/j.sbspro.2013.10.683

Hashemi, M., Azizinezhad, M., & Dravishi, S. (2012). Collocation a neglected aspect in teaching and learning EFL. *Procedia - Social and Behavioral Sciences*, *31*(2011), 522–525. https://doi.org/10.1016/j.sbspro.2011.12.097

Howarth, P. (1998). The phraseology of learners' academic writing. *Applied Linguistics*, *19(1)*, 24–44.

Jafarpour, A. A. (2013). *A Corpus-based Approach toward Teaching Collocation of Synonyms*. *3*(1), 51–60. https://doi.org/10.4304/tpls.3.1.51-60

Laufer, B. Waldman, T. (2011). Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning*, *61*(2), 647–672. https://doi.org/https://doi.org/10.1111/j.1467-9922.2010.00621.x

Lewis, M. (1993). *The lexical approach*. Hove: Language Teaching Publications.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. Annual Review ofApplied Linguistics. *Annual Review OfApplied Linguistics*, *32*, 130–149. https://doi.org/10.1017/S0267190512000098

Shammas, Dr. N. A. (2013). Collocation in English : Comprehension and Use by MA Students at Arab Universities Dr . Nafez Antonious Shammas Faculty of Arts & Sciences Petra University Amman The Hashemite Kingdom of Jordan. *International Journal of Humanities and Social Science*, *3*(9), 107–122.

Webb, L & Kagimoto, E. (2012). The Effects of Vocabulary Learning on Collocation and Meaning. *TESOL Quarterly*, *43*(1), 55–77. https://doi.org/https://doi.org/10.1002/j.1545-7249.2009.tb00227.x