

## COLLOCATIONS IN ESL WRITING: HEAD VERB FREQUENCY EFFECTS AND MALFORMED COLLOCATIONS RESPONSES

Hasliza Abd Halim

Universiti Utara Malaysia, haslieza@uum.edu.my

*Accepted date:* 21-02-2019

*Published date:* 08-04-2019

*To cite this document:* Halim, H. A. (2019). Collocations in ESL Writing: Head Verb Frequency Effects and Malformed Collocations Responses. *International Journal of Education, Psychology and Counseling*, 4(27), 91-104.

---

**Abstract:** *Collocations are words that commonly occur together or near each other in a text (Coxhead, 2006), for example, make a decision and foot the bill. Collocations and phrases are important because they help with fluency in writing and even speaking. This study explores the vocabulary knowledge of speakers of Malaysian English as it is assumed that non-native speakers of standard English do not share similar advantages to native speakers. It is due to the fact that non-native speakers, particularly adult learners, are normally expected to acquire words rather than phrases (Kuiper, Columbus & Schmitt, 2009). In addition to that, Wray (2002) claims that non-native speakers acquire individual words separately which later pair for correct collocations. Thus, this study examined the collocations acquired by Malaysian learners with exposure to local English. The study is looking at restricted verb-noun collocations of written English. The objective of the study is to assess the effect of head verb frequency on the acquisition of English restricted collocations. A group of foundation students who participated in the study have answered a set of cloze tests (Halim, 2014) and produced an essay each. The results show that there is a moderate and positive relationship between the head verb frequency and the test scores in the case high (light) frequency verbs of giving, stop make, get, and one medium frequency head verb, clear. The set of malformed collocations revealed the types of responses learners tend to come up with and indirectly illustrate the challenge the learners encounter in mastering restricted collocations. What is observed is that many of the non-idiomatic responses are from high light frequency and high frequency verbs. This suggests that the second hypothesis, that the verb choice made for the non-idiomatic answers would be at the high end of the frequency spectrum, was supported.*

**Keywords:** *Verb-Noun Collocations, Malaysian English, Head Verb Frequency, Malformed Collocations*

---

## **Introduction**

Collocations are words that commonly occur together or near each other in a text (Coxhead, 2006). Collocations as units of formulaic language are definitely regarded as one of the mediums in interpreting and shaping our understanding of language learning. There has been a notable increase in interest in this research area as demonstrated by (Alali & Schmitt, 2012; Halim, 2014; Gablasova, Brezina & McEnery, 2017; Wray, 2002; Moon, 1997, 1998; Kuiper, 2004; Koya, 2005; Howarth, 1996).

Various recent studies have paid particular attention on the acquisition of collocations either by native or non-native speakers (Paquot & Granger 2012; Durrant and Siyanova 2015; Durrant and Schmitt 2009; Nesselhauf, 2005; Nguyen&Web, 2016; Granger & Bestgen 2014). These studies provide compelling evidence that collocations are deemed important in investigating language acquisition. Plus, the advantage of having access to corpora as mediums providing authentic and rich source of data.

So, in the case of Malaysia, English language is regarded as a second language. Regardless of the mode as a second language, English is likely to continue to be important for Malaysians in world interactions. The establishment of the local variety of local English in Malaysia has become the pride of all Malaysians with its local nuances and innuendos which is reflected from the localized vocabulary, pronunciation as well as pragmatic features. Indirectly, collocational studies would be another avenue to investigate the local English varieties of vocabulary research.

Given the above, the present study embarks on a study of vocabulary acquisition. In particular, it examines the English collocations known by speakers of Malaysian English. The motivation for conducting this study is to explore the vocabulary knowledge of speakers of Malaysian English as it is assumed that non-native speakers of standard English do not share similar advantages to native speakers. It is due to the fact that non-native speakers, particularly adult learners, are normally expected to acquire words rather than phrases (Kuiper, Columbus & Schmitt, 2009). Wray (2002) claims that non-native speakers acquire individual words separately which later pair for correct collocations.

The above notion has call for an urge to examine the lexical collocations acquired by Malaysian learners with exposure to local English. The study is restricted to Verb-Noun collocations of written English. The objective of the study is to assess the effect of head verb frequency on the acquisition of English restricted collocations. In this sense, UUM Foundation students would be the right target participants as they are management students who have to equip themselves with the English language skills. The matter of exploring their language acquisition is essential as they are expected to deliver as good language users in terms of speaking and writing.

## **Objectives of The Research**

The specific objectives to be achieved involve seeking answers to the following research questions:

1. How does the frequency of the verb frequency affect the acquisition and the production of restricted collocations?
2. What are the malformed collocations responses produced by the learner/s?

The frequency of the head verbs in corpora predicts acquisition because the hypothesis is that the frequency of a lexical item is a proxy for the likelihood that a learner has been exposed to

the item. So, it is important to investigate the relationship of the head verb frequency with acquisition.

The second objective of the study is to reveal the malformed collocations produced by the learners. The patterns which later formed as a small written database would definitely be an authentic source of evidence of ESL writing materials.

This study will adopt a model of lexical access for phrasal lexical items, namely superlemma theory (Sprenger et al., 2006; Kuiper et al., 2007). This theory along with other relevant theories by Cutting and Bock (1997) and Sprenger, Levelt and Kempen (2006) look at how phrasal lexical items are stored and retrieved as well as looking at what is acquired. This framework is deemed important in explaining how retrieval from the mental lexicon takes place in cloze tests. Apart from this model, Sinclair's (1991) model of the way words occur in a text is used. This model has outlined the distinction two major concepts between the *open-choice principle* and the *idiom principle*. In this sense, the *open-choice principle* is where language text is seen as a series of choices where the only limitation on choice is grammaticalness. This principle is suggesting *slot-and-filler* model, with the idea that language is creative and operates simultaneously on several levels. Thus, a wide variety of possible words can be filled into each *slot*. This could probably be the traditional way of describing language. Whereas, the *idiom principle* proposes that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments. The idiom principle illustrates the fact that there are patterns or regularities in how words co-occur with each other.

### **Collocations and Frequency**

Collocation has come under the spotlight the establishment of most influential work done by Palmer (1933) and Firth (1957). These studies have motivated many scholars to explore the phenomenon even further. It resulted on many researches been discussed on not only on each word in a sentence, but on the combination of words in terms of productivity (Wray, 2002; Moon, 1997, 1998; Kuiper, 2004; Howarth, 1996, 1998; Gablasova, Brezina & McEney, 2017; Durrant and Schmitt 2009; Millar, 2011; Nesselhauf, 2005; Nguyen&Web, 2016)

Nation (2001) has highlighted that fluency is developed through repeated encounter of collocational sequences. So, learners need multiple exposure either explicitly or implicitly. However, native speakers can fluently say multi-clause utterances. This is due to the fact that those formulaic varieties are already memorized as prefabricated phrases. The phrases are stored as single wholes and are instantly available for use without the cognitive load of having to assemble them on-line as one speaks (Pawley and Syder, 1983; Kuiper and Haggio, 1984; Kuiper, 1996). Pawley and Syder propose that our mind uses its vast memory to store these fabricated phrases in order to compensate for limited working memory.

The above studies have inspired more research done on collocations used by either native or non-native. Studies by Howarth (1996) presents a very significant study on the use of prefabricated language in the production of native and non-native writers of English. This study has led to an establishment of a framework which focuses on restricted collocations. The findings have revealed that some deviations are found in the writing of advanced foreign language learners from that of native academic writing, and this is due to lack of knowledge of what is conventional in the use of academic, field-related collocations. Some collocational errors made by the overseas postgraduate students were spotted based on the data access from the corpus of advanced learner writing which consists of academic essays. One of the findings

is that learners do not approach the phenomenon from the same direction as native speakers. At the same time teaching materials of the course may offer very little help as they fail to recognize the nature of collocations.

Granger's (1998) study explains how the usage of foreign learners deviates from such standard norms. She compares native and non-native varieties of English with the hypothesis that learners will make less use of prefabricated language (collocations and formulae) than native speakers.

### ***The Importance of Frequency Approach in Lexical Studies***

This study looked at how and why corpus frequency is a significant matter in vocabulary acquisition. The discussion will also provide insights into the relationship between the frequency of the head verbs and their learnability. It will be suggested that frequency is a proxy measure for the likelihood of a learner being exposed to a vocabulary item, including a phrasal lexical item such as a collocation.

Corpora provide us with large collections or databases of texts from a language. Specifically, the *'insights from corpus research have revolutionized the way we view language, particularly words and their relationships with each other in context'* (Schmitt, 2000: 68). Thus, it includes looking at the relationship between frequency and collocations. Large corpora, i.e. Nation's (1990) list, are required to make such a study possible, at the same time avoiding painstaking and tedious hours of manual labour. However, in using corpora for linguistics inquiry, we need to bear in mind the cautions made by Biber's (1989) study of the difference between written and spoken corpora.

It is impossible to judge how many words individual people are exposed to as there are no records of personal corpora, i.e. corpora 'in the head' containing everything that an individual has heard or acquired. Thus, the only accessible and possible corpora are the ones containing texts of a more general kind, i.e. text corpora. Thus, in this sense corpus frequency is taken as a proxy for the probability that a language learner has been exposed to a lexical item. Frequent collocations will therefore be the most useful because *'frequent collocations have greater chances of being met and used'* (Shin and Nation, 2008). Shin and Nation also found that *'the shorter the collocation, the greater the frequency'*. This study revealed that two-word collocations make up 77 percent of the total number of collocations in the spoken section of the British National Corpus (BNC).

However, the present study is focusing only on the frequency of the head verbs and not the frequency of the collocations. This attempt been made due the fact that learners might be exposed to more individual words rather than formulaic expressions. What is more interesting is that vocabulary acquisition is generally known to be sensitive to the frequency of vocabulary items (Ellis, 2002; Kuiper, Columbus & Schmitt, 2009; Read, 1988; Schmitt, Schmitt & Clapman, 2001; Schmitt & Schmitt, 2012; Schmitt, 2010; Trembley, Baayen, Derwing & Libben, 2008; Cobb, 2007; Gass & Mackey 2002; Gonzales Fernandez & Schmitt, 2015).

The frequency boundaries tabled by Schmitt and Schmitt (2012) are quite similar to Kuiper, Columbus & Schmitt's (2009) frequency of lemmatized verbs where the three frequency bands are structured in the same way, but the frequency range is ranked into four categories, with high frequency vocabulary divided into light (or de-lexicalised) verbs (Grimshaw, 1990) and non-light high frequency verbs. The four categories are classified as shown in Table 1

**Table 1 Categorization of Verbs in Kuiper, Columbus & Schmitt (2009)**

Category	Frequency criterion
<b>High frequency light verbs (HL)</b>	Appearing in the top 1-3,000 words in the MFW lists (as words). Note that light verbs are also higher in frequency than the other high frequency verbs
<b>High frequency lexical verbs (H)</b>	Appearing in the top 1-3,000 words in the MFW lists (as words)
<b>Medium frequency lexical verbs (M)</b>	Appearing in the 3,000-5,000 word list in the MFW lists (as words)
<b>Low frequency lexical verbs</b>	Not appearing in any lists

A study by Cobb (2007) has provided significant insight supporting the frequency distribution. The aim is to see how often they occur in a 517,000-word extract of the Brown written English corpus. This notion of frequency is also addressed by Coxhead (2000, 2011) as the Academic Word List (AWL0 where this list is extensively used in English for academic purposes (EAP) classrooms.

The above literature supports the approach taken in the present study of making the relevant word frequency list by ranking the verbs in 3 frequency levels: high-frequency, medium-frequency and low-frequency. So, the present study will adopt the frequency list by Halim (2014) which particularly extracted from NST Corpus. The selection criterion of verb categorization of NST corpus is listed below.

**Table 2 Categorization of Verbs in NST Corpus**

Category	Frequency criterion
High frequency light verbs (HLF)	Appearing in the top with the highest occurrences to 20,000 occurrences in the NST corpus list (as verbs only). Note that light verbs also tend to be higher in frequency than the other high frequency verbs. (Rank number (N) 1-1000verbs)
High frequency lexical verbs (HF)	Appearing in less than 20,000 to 5,000 words in the NST corpus list (as verbs only). (Rank number (N) 1-1000verbs)
Medium frequency lexical verbs (MF)	Appearing in less than 5,000 to 200 words in the NST corpus list (as verbs only). (Rank number (N)1000-3000verbs)
Low frequency lexical verbs (LF)	1-200 occurrences in the NST corpus. (Rank number (N) more than 3000verbs)

## Methodology

### *Research Design*

The present study will employ correlation design under quantitative research to analyze the results. Logistic binary regression was also done to investigate the effects of head verb frequency on acquisition. Also, the analysis of the non-idiomatic instances was made to see the verb options made by learners. The objective was to investigate whether the verb options made are from the high frequency band. In other words, the present research project is both quantitative and qualitative in nature.

### *Instruments*

A relevant cloze test (Halim, 2014), a tagger (CLAWS Tagger), a processing software (Wordsmith Tool 6.0) and students' non-idiomatic options. At the same time verb lemma list from NST Corpus were used to check verb frequency made from malformed collocations.

### *Participants*

The participants who participated in the study were 21 candidates of term 2 UUM foundation students. From the background questionnaire, the average age of the students will be measured.

### *Data Collection Procedures*

All ethical considerations were maintained while conducting the present study. The permission of the concerned authority of the university was sought before conducting the study. The participation of the students for the tests measure will be voluntary. The researcher took all the responsibility of administering the cloze test meant for the study. The time allocated for each of the session will be strictly maintained, and the researcher will ensure that the instructions and explanations of all the tests would be clear and understandable to all participants, and they would receive the same type of instructions and explanations.

The researcher has notified the students about the overall objective of the study and inform them that the performance of them (the students) on the tests would not influence the academic achievement of the course. The data collection procedure for the present study was in the form of paper and pen testing. During the class periods, the students will get the tests and willing they will participate in the tests.

The present research is testing the following research questions:

**Table 3 Research Questions**

	<b>Research Questions</b>	<b>Tools</b>	<b>Analysis</b>
<b>RQ 1</b>	1.How does the frequency of the head verbs of restricted collocations affect the acquisition and the production of English collocations?	Cloze Test (Hasliza, 2014)	Correlation (SPSS)
<b>RQ 2</b>	2.What are malformed collocations produced by the learners?	Non-idiomatic responses,	NST lemma list

## Results and Discussions

### *Study 1*

The respondents were asked to fill lexical verb gaps from a text written by the researcher in a vernacular style. It was about a multicultural event celebrated in Malaysia. The aim was to maintain stylistic homogeneity throughout the task and provided sufficient narrative interest to encourage respondents to maintain their interest until the end of the story (Kuiper, Columbus, & Schmitt, 2009). Given below are the 20 restricted collocations tested on the students. The students were only required to provide the missing verbs from the given text.

**Table 4 The Frequency of Head Verbs and The Frequency of Restricted Collocations in NST Corpus**

Restricted collocations	Frequency band	Total occurrences (of head)	Frequency of (the exact) PLIs in NST corpus
does wonders	HL1	61 462	32
make a fast buck	HL2	37 075	20
taking a big risk	HL3	33 377	143
get a grip of oneself	HL4	23 780	6
give a hoot	HL5	20 504	3
look the part	H1	14 222	12
tell the difference	H2	10 770	27
pay respect	H3	6 789	99
create a win-win situation	H4	5 218	6
stop bickering	H5	4 261	7
kill time	M1	1 997	13
steal the show	M2	913	36
cleared backlog	M3	907	28
observe taboo and prohibited things	M4	728	2
air view	M5	450	31
crack(a) joke	L1	186	16
shouldered the responsibility	L2	115	39
rekindle family ties	L3	55	7
foot the bill	L4	22	27
gnash teeth	L5	3	2

*Notes:*

HL- High Light frequency

H- High frequency

M- Medium frequency

L- Low frequency

The following table shows the results of correlation index for the relationship between the head verb frequency and the test score. The Spearman correlation results indicate the strength for each of the 20 head verbs, from high light frequency, high frequency, medium and low frequency.

**Table 5 Correlation Results**

Correlations			HL5	H4	H3	M2	HL2	HL4	M4	L1	HL3	Score	
Spearman's rho	HL5	Correlation Coefficient	1.000	.499*	.141	.447*	-.043	-.030	.141	.141	-.344	.591**	
		Sig. (2-tailed)		.021	.541	.042	.853	.897	.541	.541	.126	.005	
		N	21	21	21	21	21	21	21	21	21	21	
	H4	Correlation Coefficient	.499*	1.000	.108	.086	.198	.462*	.108	.108	-	.495*	.639**
		Sig. (2-tailed)	.021		.640	.712	.390	.035	.640	.640	.022	.002	
		N	21	21	21	21	21	21	21	21	21	21	
	H3	Correlation Coefficient	.141	.108	1.000	-.158	-.091	.235	-.050	-.050	-.091	.247	
		Sig. (2-tailed)	.541	.640		.494	.694	.306	.830	.830	.694	.280	
		N	21	21	21	21	21	21	21	21	21	21	
	M2	Correlation Coefficient	.447*	.086	-.158	1.000	.289	.135	.316	.316	-.289	.627**	
		Sig. (2-tailed)	.042	.712	.494		.204	.560	.163	.163	.204	.002	
		N	21	21	21	21	21	21	21	21	21	21	
	HL2	Correlation Coefficient	-.043	.198	-.091	.289	1.000	.156	.548*	.548*	-.167	.486*	
		Sig. (2-tailed)	.853	.390	.694	.204		.500	.010	.010	.470	.026	
		N	21	21	21	21	21	21	21	21	21	21	
HL4	Correlation Coefficient	-.030	.462*	.235	.135	.156	1.000	-.213	-.213	-.389	.535*		
	Sig. (2-tailed)	.897	.035	.306	.560	.500		.353	.353	.081	.012		
	N	21	21	21	21	21	21	21	21	21	21		

M4	Correlation Coefficient	.141	.108	-.050	.316	.548*	-.213	1.000	1.000**	-.091	.380
	Sig. (2-tailed)	.541	.640	.830	.163	.010	.353			.694	.089
	N	21	21	21	21	21	21	21	21	21	21
L1	Correlation Coefficient	.141	.108	-.050	.316	.548*	-.213	1.000**	1.000	-.091	.380
	Sig. (2-tailed)	.541	.640	.830	.163	.010	.353			.694	.089
	N	21	21	21	21	21	21	21	21	21	21
HL3	Correlation Coefficient	-.344	-.495*	-.091	-.289	-.167	-.389	-.091	-.091	1.000	-.347
	Sig. (2-tailed)	.126	.022	.694	.204	.470	.081	.694	.694		.123
	N	21	21	21	21	21	21	21	21	21	21
<b>Score</b>	<b>Correlation Coefficient</b>	<b>.591**</b>	<b>.639**</b>	<b>.247</b>	<b>.627**</b>	<b>.486*</b>	<b>.535*</b>	<b>.380</b>	<b>.380</b>	<b>-.347</b>	<b>1.000</b>
	Sig. (2-tailed)	.005	.002	.280	.002	.026	.012	.089	.089	.123	
	N	21	21	21	21	21	21	21	21	21	21

\*. Correlation is significant at the 0.05 level (2-tailed).

\*\*. Correlation is significant at the 0.01 level (2-tailed).

There are moderate and positive relationship between the head verb frequency and the test scores. As shown in the correlation table above, the correlation index for the relationship between the high light frequency verb (HL5) and the test score is 0.591, which is between 0.4-0.7. The correlation index for the relationship between the high frequency verb (H4) and the test score is 0.639, which is between 0.4-0.7. The correlation index for the relationship between the medium frequency verb (M2) and the test score is 0.627, which is between 0.4-0.7. The correlation index for the relationship between the high light frequency verb (HL2) and the test score is 0.486, which is between 0.4-0.7. The correlation index for the relationship between the high light frequency verb (HL4) and the test score is 0.535, which is between 0.4-0.7.

So, the results from these analyses indicate that there is a moderate, positive relationship between certain head verb frequencies. And those head verbs are mostly from either high light frequency or high frequency verbs, 4 from high light verbs, and 2 from high frequency verbs. Basically, the idiomatic instances made are from the high frequency band head verbs. So, this has proven that the frequency of the head verbs in corpora predicts acquisition as the frequency of a lexical item is a proxy for the likelihood that a learner has been exposed to the item. So, in

this case the acquisition is only related to the high frequency verbs GIVE, STOP, MAKE, GET, and one medium frequency head verb, CLEAR.

## Study 2

### *Malformed Collocations*

Among the 21 participants, only 1 learner's profile is presented as a means of providing detailed documentation of an individual's personal lexical knowledge based on the cloze test results. This learner's non-idiomatic responses were listed and coded and also checked for verb frequency rank with NST corpus. The motivation for proceeding with this analysis was to test the second hypothesis with the presumption of the use of more verbs at the high frequency end of the spectrum for other non-idiomatic verbs.

### *A Case Study - Samples of an Individual's Set of Responses*

In this section a student's answers are presented and discussed in detail. Umi's non-idiomatic responses were analyzed using the mean results. Umi's (not her real name) answers were all non-idiomatic. The following table shows Umi's responses for non-idiomatic answers.

**Table 6 Student 5 (T5) –Umi's Responses**

Idiomatic verb	Student's answer/verb	Frequency in NST corpus	Head verb band
1.shoulder(ed)	HAD	194,766	<b>High frequency</b>
2.give	WANT	17,072	<b>High frequency</b>
3.stop	DO (not)	49,762	<b>High frequency</b>
4.pay	GIVE	21,064	<b>High frequency</b>
5.observed	AVOID	2,081	
6.rekindle	BOND	703	
7.kill	FULFILL	845	
8.gnash	SHOWN	13,832	<b>High frequency</b>
9.air	POINT	8,306	<b>High frequency</b>
10.clear(ed)	FINISHED	4,126	
11.tell	SEE	18,411	<b>High frequency</b>
12.make(making)	GETTING	24,101	<b>High frequency</b>
13.get	HAVE	194,766	<b>High frequency</b>
14.does	NO	-	
15.steal	DISTURB	-	
16.look	IS	190,596	<b>High frequency</b>
17.crack	MAKE	39,360	<b>High frequency</b>

18.foot	PAY	7,640	<b>High frequency</b>
19.take (taking)	IN	-	
20.create	LOVE	6,969	<b>High frequency</b>
<b>Total</b>			<b>13</b>

Umi's score for the idiomatic responses was 0%. The non-idiomatic verbs were checked using a frequency list from NST corpus. The verbs been analysed based on the verb lemma list in NST corpus where they been categorized into 4 bands (Table 2). The above table shows that 13 verbs are form high frequency verbs, with the appearance of more than 2000 verbs in the corpus. The results reveal that the verb choices made by Umi were highly frequent and could be categorized within the highest verb frequency category. This suggests that the second hypothesis, that the verb choice made for the non-idiomatic answers would be at the high end of the frequency spectrum, was supported.

### Conclusion

The results from these analyses indicate that there is a moderate, positive relationship between certain head verb frequencies, this has proven that the frequency of the head verbs in corpora predicts acquisition as the frequency of a lexical item is a proxy for the likelihood that a learner has been exposed to the item. So, in this case the acquisition is related to the high frequency verbs GIVE, STOP, MAKE, and GET, which are among the high ranked verbs. The results has proven the notion of non-native learners and native speakers, where non-native speakers of standard English do not share similar advantages to native speakers. It is due to the fact that non-native speakers, particularly adult learners, are normally expected to acquire words rather than phrases (Kuiper, Columbus & Schmitt, 2009). Wray (2002) claims that non-native speakers acquire individual words separately which later pair for correct collocations.

The findings of study 2 are significant because they illustrate the types of responses learners tend to come up with and indirectly illustrate the challenge of mastering restricted collocations. In this sense, the malformed or infelicitous restricted collocational choices made by L2 learners may reflect the struggle of learners learning a learning.

This study reinforces that there are challenges of mastering restricted collocations and the use of these sequences. May be within the same speech community collocations are less used and expected, but for academic writing and purposes, it has been generally agreed that the appropriate use of these sequences is highly required (Li and Schmitt, 2009).

The results have shown on how collocations been retrieved as in the model of lexical access for phrasal lexical items, namely superlemma theory (Sprenger et al., 2006; Kuiper et al., 2007). It looked at how phrasal lexical items are stored and retrieved as well as looking at what is acquired. This theoretical framework has helped in explaining how retrieval from the mental lexicon takes place in cloze tests.

The non-idiomatic instances made by learners revealed and supported Sinclair's principle of *slot-and-filler* model, with the idea that language is creative and operates simultaneously on several levels. Thus, a wide variety of possible words can be filled into each *slot*. So, if learners like Ummi do not have enough collocations in their mental lexicon, native-like competency is hindered, requiring her to opt for other strategies when having language difficulties. A related strategy which might be applied by the participants is *guessing from context* which is

commonly used for learning individual words (Nation, 1990, 2001). The study by Millar (2011) has shown that malformed L2 collocations lead to an increased processing burden for native speakers in terms of slower reading speed. However, some of the same receptive processing effects could also be hypothesized for other aspects of language use. At the same time, it is also suggested that malformed collocations should be viewed in a positive way (Halim and Kuiper, 2018)

Since collocations are vital in language learning there might be various opinion on how collocations are best learnt. Schmitt (2000) has raised the issue of how language learners are able to acquire thousands of word families. Schmitt argued that this amount is probably too large to be learnt solely from formal study, so collocational knowledge is best acquired implicitly, through extensive exposure to the target language. Nation (2001) further suggests that fluency is developed through repeated encounter of collocational sequences. So, learners need multiple exposure either explicitly or implicitly.

## References

- Alali, Fatima A. & Schmitt, N (2012) Teaching formulaic sequences: the same as or different from teaching single words? *TESOL Journal* p.153-180
- Biber, D. (1989). A typology of English texts. *Linguistics*, 27, 3-43.
- Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning and Technology*. 11.3, 38-63.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly* 34.2, 213-238
- Coxhead, A. (2011). A new academic word list 10 years on: research and teaching implications. *TESOL Quarterly* 45.2, 355-362
- Cutting, J. and Bock, K. (1997). That's the way the cookie bounces: syntactic and semantic components of experimentally controlled idiom blends. *Memory and Cognition* 25: 1, 57-71.
- Durrant, P., and Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics*, 47(2), 157-177.
- Durrant, P., & Siyanova-Chanturia, A. (2015). Learner corpora and psycholinguistics. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 57-77). Cambridge, UK: Cambridge University Press.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143-188.
- Firth, J. R. (1957). 'A Synopsis of Linguistic Theory, 1930-1955' In G. Kennedy (Ed.), *An Introduction to Corpus Linguistics*. New York: Addison Wesley Longman, Inc.
- Gablasova, D., Brezina, V. & McEnery, T. (2017). Collocations in corpus-based language learning research: identifying, comparing and interpreting the evidence. *Language Learning*, 67, 155-179.
- Gass, S. M., & Mackey, A. (2002). Frequency effects and second language acquisition. A complex picture? *Studies in Second Language Acquisition*, 24, 249-260.
- González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? The effects of frequency and amount of exposure
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing; Collocations and formulae. In A. P. Cowie (ed.), *Phraseology: Theory, analysis and applications*. (pp. 185-204) Oxford: Clarendon Press.

- Granger, S., & Bestgen, Y. (2014). The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. *International Review of Applied Linguistics in Language Teaching*, 52, 229–252.
- Grimshaw, J. (1990). *Argument Structure*. Cambridge, MA: MIT Press.
- Hasliza Abd Halim and Kuiper, K. (2018). Individual differences in the acquisition of restricted collocations. *International Journal of Education, Psychology and Counseling*, 3, 36-49.
- Hasliza Abd Halim. (2014). *Restricted verb phrase collocations in standard and learner Malaysian English*. (Unpublished doctoral dissertation). The University of Canterbury, New Zealand.
- Howarth, P.A. (1996). *Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making*. Tübingen: Max Niemeyer Verlag.
- Koya, T. (2005). *The Acquisition of Basic Collocations by Japanese Learners of English*. Unpublished doctoral dissertation. Waseda University.
- Kuiper, K. (2004). Formulaic performance in conventionalised varieties of speech. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing and Use* (pp. 37-54). Amsterdam: John Benjamins Publishing Company.
- Kuiper, K., Columbus, G., & Schmitt, N. (2009). Acquiring phrasal vocabulary. In S. Foster-Cohen (Ed.), *Advances in Language Acquisition*. Basingstoke: Palgrave Macmillan.
- Kuiper, K. and Haggio, D. C. (1984). Livestocks auctions, oral poetry and ordinary language. *Language in Society*, 13, 205-34.
- Kuiper, K. (1996). *Smooth Talkers*. Mahwah, NJ: Erlbaum.
- Kuiper, K., van Egmond, M-E., Kempen, G. M. and Sprenger, S. (2007). Slipping on superlemmas: multiword lexical items in speech production. *The Mental Lexicon*, 2: 3, 313-57.
- Li, J., and Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, 18, 85-102.
- Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics*, 32, 129-148.
- Moon, R. (1997). Vocabulary connections: Multi-word items in English. In N. Schmitt and M. McCarty (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 40-63), Cambridge: Cambridge University Press.
- Moon, R. (1998). *Fixed Expressions and Idioms in English: A Corpus Based Approach*. Oxford: Clarendon Press.
- Nation I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nesselhauf, N. (2005). *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.
- Nguyen, T. M. H., & Webb, S. (2016). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, 1–23.
- Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, 32, 130–149.
- Palmer, H. E. (1933). Second Interim Report on English Collocation. In G. Kennedy (Ed.), *An Introduction to Corpus Linguistics*. New York: Addison Wesley Longman, Inc. Tokyo: Institute for Research in English Teaching.
- Pawley, A. and Syder, F. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. Richards and R. Schmidt (eds), *Language and Communication* (pp. 191-226). London: Longman.
- Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal* 19, 12-25.

- Schmitt, N., (2000) *Vocabulary in Language Teaching*. Cambridge University Press. United Kingdom.
- Schmitt, N., Schmitt, D., and Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55-88.
- Schmitt, N. (2010). *Researching Vocabulary-A Vocabulary Research Manual*. London: Palgrave Macmillan.
- Schmitt, N and Schmitt, D. (2012) A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 4, 484-503.
- Shin, D. and Nation, P. (2008). Beyond single words: the most frequent collocations in spoken English. *ELT Journal*, 62/4: 339-348.
- Sinclair, J. (1991). *Corpus, Concordance and Collocation*. Oxford: OUP.
- Sprenger, S., Levelt, W. J. M. and Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, 54, 161–84.
- Tremblay, A., Bayern, R. H., Derwing, B., and Libben, G. (2008). *Frequency and the processing of multiword strings: A behavioural and ERP study*. Paper presented at the conference of the Canadian Linguistics Association, Université de la Colombie-Britannique, May 31 2008. <http://ocs.sfu.ca/fedcan/index.php/cla/acl-cla2008/paper/viewFile/219/152>
- Wray, A. (2002), *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.