Volume: 3 Issues: 16 [September, 2018] pp.36-49]
International Journal of Education, Psychology and Counseling
eISSN: 0128-164X
Journal website: www.ijepc.com

INDIVIDUAL DIFFERENCES IN THE ACQUISITION OF RESTRICTED COLLOCATIONS

Hasliza Abd Halim [1]
Koenraad Kuiper [2]

[1]Universiti Utara Malaysia, haslieza@uum.edu.my
[2]University of Canterbury, konkuiper@gmail.com

Accepted date: 31 October 2017 **Published date**: 13 September 2018

To cite this document: Halim, H. A., & Kuiper, K. (2018). Individual Differences in The Acquisition of Restricted Collocations. *International Journal of Education, Psychology and Counseling*, *3*(16), 36-49.

Abstract: *The acquisition of verb-noun collocations (e.g. make a mistake) causes great difficulties to (adult) L2 learners for several reasons (Boers, Demecheleer, Coxhead and Webb, 2014). Thus, investigating the use of collocation in English language learning is important as such study may inform us on the use of restricted collocations in English language teaching and learning including in the Malaysian context. Apart from the difficulties in the acquisition of collocations, Dornyei and Skehan (2003), and Sawyer and Ranta (2001), have found that individual difference factors have significant impacts on language learning in general. Apart from that, Schmitt, Dornyei, Adolphs & Durow (2004) argue that these individual differences might also influence the acquisition of formulaic language. The results of this study provide support for the above finding. Individual differences are indeed a factor. A new testing approach is proposed; the semantic plausibility metric, which is used as a tool for this study, and is shown to be useful as a measure of vocabulary acquisition as well as for looking at learners' test taking strategies (Halim, 2014). This study also suggests that malformed collocational choices should be viewed positively.*

Keywords: *Restricted Collocations, Individual Difference, Semantic Plausibility Metric, Malformed Collocations*

Introduction

Restricted collocations appear in all types of speech and can be defined as *'pairs of words which occur together in ways that are more restrictive than the grammar of the language requires'* (Kuiper, 2004: 51). Restricted collocations are not formulae as they are not restricted by anything accept for their meaning. Kuiper uses *give offence* and *take offence* as examples. The only 'acceptable' verbs used by native speakers of English in this collocational context are 'restricted' to these two verbs. It is impossible to use *donate offence* or *accept offence*.

Apart from Kuiper, Howarth (1996) describes research done on phraseological performance of non-native writers of English in academic writing, in which the findings are significant for the study of collocations. Howarth's definition of restricted collocation is as follows:

> ‘combinations in which one component is used in its literal meaning, while the other is used in a specialized sense. The specialized meaning of one element can be figurative, delexical or in some way technical and is an important determinant of limited collocability at the other..’
>
> (Howarth,1996: 47)

In defining restricted collocation, Cowie (1991: 102) describes restricted collocations as ‘word-combinations in which one element (usually the verb) has a technical sense, or a long-established figurative sense which has since lost most of its analogical force’. Cowie discusses a few examples such as *run a deficit*, *abandon a principle*, or *champion a cause*, in which the object noun limits the choice of verb to only one or two. Moon (1998: 27) sees that this kind of collocation occurs where ‘a word requires association with a member of a certain class or category of item’. Moon further proposes that they are semantically and lexicogrammatically restricted. Moon shares Aisenstadt’s (1981) concept of ‘restrictedness’ by stating that a word contains a particular meaning only when it is collocated with certain other words. Aisenstadt (1981) refers to these occurrences as restricted collocations and provides the examples of *face the truth/ facts/ problems*.

In relation to that, Granger (1998) investigates restricted collocations which focus more on amplifiers functioning as modifiers of adjectives. The findings show ‘sharp differences between native and non-native usage’ (Cowie, 1998: 13). Granger’s study finds that *completely* and *totally* are significantly overused by the learners. According to Kuiper (2009) restricted collocations involve preferential selection of word combinations where the combinations are arbitrary, and they might also be idiomatic.

Moon (1998) found that verb phrase idioms are the most frequent fixed expressions in the Hector Corpus. Cowie (1992) also reports the percentage of verb phrase idioms and restricted collocations in news stories and feature articles to be around 40 percent. Verb-noun combinations are regarded as key combinations in producing clauses and sentences, and they are the most often selected in previous empirical research (e.g. Bahn and Eldaw, 1993; Bahns, 1993; Biskup, 1992; Nesselhauf, 2003). These studies have suggested that more focus is to be placed on verb-noun collocations, since it is the verb that causes the greatest difficulties for learners.

The acquisition of verb-noun collocations (e.g. *make a mistake*) causes great difficulties to (adult) L2 learners for several reasons (Boers, Demecheleer, Coxhead and Webb, 2014). It is typically found that learners tend to substitute the restricted verb in the collocations by a conventional choice such as e.g. *do a mistake*. A study by Laufer and Waldman (2011) found that there were hardly any differences in productive knowledge of verb-noun collocations between lower and upper intermediate groups of EFL learners. And for the substitution case it is more likely due to the interference from the mother tongue (Yamashita and Jiang, 2010; Nesselhauf, 2005). Apart from that, learners may not possibly see the need to attend to the verb in interpreting the phrase. Learners may find the verb contributes relatively little to the semantics of some collocations. The slow uptake of verb-noun collocations by learners may lie in the lack of distinctiveness of the verbs, where the verbs may be treated as synonyms by the learners.

The Objective of the Study

This study is a study of vocabulary acquisition. It examines the English collocations known by speakers of Malaysian English. The objective for conducting the 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. This is because non-native speakers, particularly adult learners, are normally presumed 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 native-speaker collocations.

The above notion calls for an examination of the lexical collocations acquired by Malaysian learners with exposure to both Malaysian English and New Zealand English. The study is restricted to Verb-Noun collocations of written English. The objective of the study is to assess learners' selectional patterns favouring particular heads of phrases. It is supposed that the production task of filling a cloze gap requires learners to fill the gaps of the test items with either restricted collocations or with semantically plausible verbs. The task of supplying the missing verbs leads to retrieving them from the mental lexicon and it is assumed that context and some of the constituents can activate the missing verbs from the mental lexicon (Jackendoff, 1995). This study outlines an approach to acquisition which is focussed on individual acquisition.

The results will propose that the findings of the research may highlight a new dimension of understanding collocational learning in Malaysia, as well as looking at the impact of collocations which are deemed malformed in native usage.

Sawyer and Ranta (2001), have highlighted the finding that individual difference factors have been shown to have significant impacts on language learning in general. Apart from that, Schmitt, Dornyei, Adolphs & Durow (2004) argue that these individual differences might also influence the acquisition of formulaic language. The results of this study provide support for the above finding. Individual difference is indeed a factor.

In the Asian context, several studies have been done in this area. A study by Kamariah Yunus and Su'ad Awab (2011) highlights the collocational competence among law undergraduates who are studying at a local university in Malaysia. A study on phrasal verbs (PVs) among Malaysian learners of English done by Rafidah Kamarudin (2013) examined the level of understanding and use of PVs. Rafidah's study was performed by survey or questionnaire independent of corpus work. Teachers' and learners' feedback were used for data collection. The associated corpus work was based on an existing corpus, English of Malaysian Students (EMAS). The overall research looked at the understanding, perception of PVs, problems faced by learners, and how PVs were used in teaching materials. Several studies have been done in Japan (Koya, 2004, 2005, 2006) looking at the acquisition of English collocations by Japanese learners. Miyakoshi (2009) conducted a study specifically on ESL learners' collocations. Her study focused on the verb-noun collocations by Japanese learners of English.

Numerous studies have been conducted on how helpful the knowledge of collocations (and formulaic language) is in second language learning in other countries. Bahn and Eldaw (1993) conducted an experiment consisting of a translation task and a gap-filling task with advanced learners of English who had German as a native language. Granger (1998) analysed the written performance of Advanced French students and found that learners overused very frequent collocations but underused creative constructions. Biskup (1992) collected interference errors made by Polish and German learners of English. There is more evidence of recent research done in the area of formulaic language in general. Studies by Laufer and Waldman (2011) concerning verb-noun collocations, Siyanova and Schmitt (2007) on phrasal and prepositional verbs, and Millar (2011) on the impact of malformed collocations are among the related studies.

Lindstromberg and Boers (2008a, 2008b) examined the mnemonic benefits of drawing learners' attention to sound repetition commonly manifested in formulaic sequences, namely, alliteration and assonance.

Although English has had the status of a second language in the Malaysian education system for decades, many Malaysian learners are regarded as error prone in their use of English (Marlyna Maros, Tan & Salehuddin (2007); Saadiyah Darus & Kaladevi, 2009). Consequently, it is worth investigating factors causing those 'errors' or non-native-like expressions made by learners. Along the same lines, research by Ang et al. (2011) highlighted that the acquisition of phrasal vocabulary, specifically verb-phrase restricted collocations can be assessed using an error analysis approach, assessing 'errors' made by learners. However, for the analysis, they applied standard British English norms. Yet, what is ideal is that the norms for second language acquisition be a combination of aspirational norms, those that the person learning the language aspires to, and actual norms, those of a target speaker community. Aspirational norms can also be those of teachers, i.e. the norms that they wish their students to aspire to. Yet, all aspirational norms are value laden. As such they can be judged as to how realistic they are.

The above discussion has suggested that the proposed research will make a useful contribution to the limited research done on the acquisition of restricted collocations in the Malaysian context. The researchers feel that by looking at the collocational patterns of Malaysian students' performance, the features and patterns of learners' collocations can been identified.

For this study, investigating collocational acquisition in English language learning is important as such study may inform us on the use of collocations in English language teaching in Malaysian context and the local school syllabus. It is intended that the findings of the research may provide knowledge of collocations used locally. It will do this by identifying the patterns of collocations used by Malaysian English learners. Thus, this study may lead to a better understanding of the nature of acquisition of collocational patterns of written Malaysian English. Furthermore, the findings will shed light on the local norms for second language acquisition.

This can be achieved by analyzing the verb frequency list extracted from the Malaysian English corpus and the BNC, as well as investigating the learners' scores in a cloze test. This study will also suggest that relativised norms are more realistic. A new approach to assess non-native like responses is suggested by the coding for the non-native like responses being further labeled as 'semantically plausible answers' which are coded using a novel approach which seems to be more realistic. It is suggested that this is how one might go about assessing restricted collocations within the context of Malaysian second language learners learning English. Apart from that, the cloze test instrument devised for this study and the corpus which has been developed for it should also prove useful tools in assessing 'errors' or non-native like restricted collocations of Malaysian learners.

Literature Review

Superlemma Theory

This study adopts 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 theoretical framework is necessary in explaining how retrieval from the mental lexicon takes place in cloze tests.

This model merges the model of idiom production (Levelt, 1989; Levelt & Meyer, 2000) into a contemporary model of lexical access. Levelt (1989) proposes that the mental lexicon is the organization of lexical knowledge in the mind and it allows access to various types of linguistic information at different stages of the speech production process. Levelt further asserts that the mental lexicon consists of interconnected nodes that encode lexical information at various levels of abstraction. Superlemma theory assumes that a phrasal lexical item has a single lexical concept. The superlemma is the unitary representation of a lexical phrase which consists of constituent lemmas of the idiom and their unique syntactic properties. In other words, formulaic sequences are stored holistically. So, what happens during speech production is that when a single lexical concept is activated, then its superlemma node is activated. The activation of the superlemma node in turn activates the lemma nodes of all its constituent words. So, once the superlemma is sufficiently activated, a user or learner may be able to retrieve the missing word to fill up the gaps in cloze test. This process provides evidence that cloze testing is an appropriate method for investigating the acquired phrasal lexical items. In other words, if a language user is able to provide the missing word, he or she is assumed to have the knowledge of that particular expression. The process of retrieving the missing words involves a move from perception to production since it requires a user to fill up the slot with a selected word.

Open Choice Principle* and *The Idiom Principle

This study also adopts Sinclair's (1991) model of the way words occur in a text. Sinclair has outlined the distinction between the *open-choice principle* and *the idiom principle.* The *open-choice principle* is where language text is seen as a series of choices where the only limitation on choice is grammaticalness (pp. 109). This principle is often referred as the 'slot-and-filler' model with the idea that language is creative and operates simultaneously on several levels. Therefore, a wide variety of possible words can be filled into each 'slot'. Sinclair claims that this could probably be the traditional way of describing language. 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' (p. 110*).* The idiom principle illustrates the fact that there are patterns or regularities in how words co-occur with each other. Within this view, recalling the earlier discussion, collocation is defined as the occurrence of two or more words within a short space of each other in a text (pp. 170). The pervasive nature of the idiom principle is significant enough to highlight the importance of collocation. This also suggests that there may even be a larger number of phrasal items as compared to individual words, as the idiom principle is argued to be dominant over the open-choice principle.

For this research, a task was designed for learners to retrieve from the mental lexicon either an idiomatic filler or, if that was not known, a semantically plausible verb which fits in the slot. This sort of gap filling task is not primarily a perception task but rather concerns production. If this production task induces learners to fill the gap with a unique item in terms of native speaker norms such as those in restricted collocations, rather than a semantically plausible word, then this is evidence of the phrasal lexical item being accessed as a whole from the speaker's mental lexicon. Thus, this situation is closely related to the theory of spreading activation. In speech production theory, word retrieval requires selecting a lemma, a lexical representation that is semantically and syntactically specified, from all other lemmas stored in one's mental lexicon, followed by phonological encoding of that lemma (Dell, 1986; Levelt, 1989). So, in this case, a word is selected if its activation exceeds some threshold, otherwise the most activated word is opted for after a fixed period of time (Dell, 1986). However, if that filler has not had sufficient activation in the past as part of a phrasal lexical item, or in other words there has been insufficient

exposure to it in the past, then the learners need to find something that fills the gap plausibly. If they were very uncertain in their language knowledge, then they might just put anything that comes into their head into the slot. What comes into their head must then be the result of word associations (Fitzpatrick & Izura, 2011).

Native and Non-Native Collocational Knowledge

The studies done by Howarth (1996) and Granger (1998) looked at the use of collocational sequences of native and non-native writers of English. Both studies found significant deviations from standard native norms made by the learners and, on top of that, suggested that learners did not approach the phenomenon from the same directions as native speakers. The deviations from such standard norms were traced based on the errors made as well as the fact that such sequences were used less frequently by learners. While learners may avoid using these collocational sequences due to lack of knowledge, native speakers may see this as *shortcuts* (Peters, 1983: 82; Hickey, 1993: 29; Wray, 2002: 106), as native speakers have been exposed to many of these sequences in their input since they were young (Wray, 2002). This has proven that though collocations are highly significant in communications for both learners and native speakers, apparently, they are closely related to extensive exposure to a target language which in this case is English. While learners may struggle, native speakers can fluently produce multi-clause utterances because many constituents of them are memorized as prefabricated phrases (Pawley & Syder, 1983).

This fact led to the decision of recruiting 10 native speakers for the coding purpose. They rated the learners' responses based on their acceptability and plausibility.

Methodology

This study is an exploratory study examining learners' selectional patterns favouring particular heads of phrases. It is supposed that the production task of filling a cloze gap requires learners to fill the gaps of the testing items with either restricted collocations or with semantically plausible verbs. The task of supplying the missing verbs leads to retrieving them from the mental lexicon and it is assumed that context and some of the constituents can activate the missing verbs from the mental lexicon (Jackendoff, 1995). Thus, this study will present a case study where learners' exhibit preferences for head verbs by filling in cloze gaps in a cloze test.

Cloze testing is considered one of the most suitable tools in assessing language ability, namely second language proficiency and reading comprehension skills (Alderson, 1979; Abraham & Chapelle, 1992; Dörnyei & Katona, 1993; Kobayashi, 2002). The cloze tests designed for this study consist of twenty cloze gaps in a vernacular narrative to be filled by the participants. The results were analyzed in two different binary methods. The first binary analysis supposes that the selection of the cloze item is either 'right' or 'wrong', 'native like' or 'non-native like'. The second analysis is in terms of the respondent gap-fills in terms of providing:

i) an idiomatic response
ii) a semantically plausible response

This study outlines an approach to acquisition which is focussed on individual acquisition. The coding was not binary. Elsewhere it is assumed that a respondent either knew the restricted collocation or they did not. In this study the coding is different: the respondent is presumed to know the idiomatic restricted collocation or, if they do not, they enter an alternative which makes good sense in the context to a degree, ranked on a Likert scale from 1 to 6. Less plausible responses might be, for example, either entering the wrong syntactic category, i.e. a non-verb,

or entering a verb that does not suit the context. It suggests that the respondent either does not understand the context sufficiently to enter an appropriate verb or their vocabulary knowledge is so slight that they do not know a verb which would fit the context.

The objective for proceeding with an analysis in this way is that one might suppose that the acquisition of phrasal vocabulary occurs after earlier single word vocabulary is acquired, and that it is based on more extensive exposure to the target language. So, we might suppose that a respondent who is faced with a lexical retrieval task, which is essentially the nature of a cloze test, but who does not understand the context either semantically or syntactically, will opt for a wild guess. A more advanced learner who does understand the context and has acquired sufficient vocabulary will be able to fill the cloze gap with a semantically and syntactically appropriate filler. Finally, as learners become more proficient in the target language, they will more often know and enter an idiomatic filler.

To assist with the recoding of the range of cloze test responses, 10 native speakers were assigned to rate the full range of individual responses. These native speakers of English evaluated the responses and coded them based on their acceptability in the context. Their discretion was vital to assist with the recoding of the range of cloze test responses. These native speakers' evaluation of acceptability was seen as unique, as they treat phrasal lexical items quite differently from non-native speakers.

This study will also then show the range of individual differences in each member of the cohort. The hypotheses to be investigated here are that:

1. Individuals who have higher scores of idiomatic responses will also have higher rates of plausibility for their non-idiomatic responses.
2. Of the non-idiomatic responses more will be at the high end of the lexical frequency spectrum.

Individual cloze test items can also be evaluated using the above outlined strategy for how they discriminate amongst respondents by looking at how the analysis allows for discrimination among the cloze items themselves. Some items may be 'harder' both in terms of how many idiomatic responses they elicit as well as how the non-idiomatic responses rank in plausibility. Others may be 'easier' in terms of eliciting idiomatic responses, i.e. the idiom is well known, but if it is not, then the responses may be at the less plausible end of the plausibility scale. This suggests that cloze testing can be made sensitive to individuals and that the careful choice of cloze items can elicit better evidence of vocabulary learning than just coding responses as idiomatic or non-idiomatic.

Such an approach is congruent with the work of Dornyei and Skehan (2003), as well as Sawyer and Ranta (2001), who highlight the finding that individual difference factors have been shown to have significant impact on language learning in general. In addition, Schmitt, Dornyei, Adolphs & Durow (2004) argue that it is quite logical that they might also influence the acquisition of formulaic language.

Apart from supporting the second hypothesis, a learner's profile is presented in detail here, as a means of demonstrating the opportunities for future research that combines single case studies with broader cohort profiles.

Participants

The participants were 20 final year Malaysian undergraduates from the University of Canterbury as mentioned earlier, they shared the same background; these were final year students, ranging from 17 to 23 years old, and they were doing various courses, e.g. Engineering and Geography. Before the data for this study were collected, they passed IELTS (International English Language Testing System) with a minimum of Band 6. This test is an international standardised test of English language proficiency. It is jointly managed by the University of Cambridge ESOL Examinations and the British Council. The participants' results were used as a prerequisite to studying in New Zealand. Participants' native languages were Malay, Tamil, Mandarin, Cantonese or other Chinese languages. They had been learning English in school from age 7, and had at least 11 to 17 years of overall exposure to English. These students were sharing accommodation with other Malaysian friends since there is a reasonably large Malaysian student community in the university campus.

Among the 20 participants, one 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 Kilgarriff's lemmatized BNC frequency list (Reference). The motivation for proceeding with this analysis was to test the second hypothesis with the presumption of the use of more verbs at the low frequency end of the spectrum if less exposure was received.

Procedure

As mentioned earlier, ten native speakers were asked to assess the cloze test responses. They were asked, 'Does the insertion of each of the following words into the story at this point make good sense or not?' Their task was to place a score of 1-6 in the provided column by indicating how acceptable they found the word in the given context.

Their scores:

1- completely unacceptable- I can't imagine this word being used in this context
2-
3-
4-
5-
6- entirely acceptable – I would use this word in this context

They were informed that the most obvious answer may be missing from the list and they were required to rate each word as it fitted in the gap. The mean value and standard deviation were then calculated across the responses in the two new categories. Provided below is the list of the verbs used for the analysis. Note that the verbs are specific to each cloze gap. The phrasal lexical items used in the cloze test are shown in Table 1. They are classified according to their relative frequency in the BNC as high light (HLF), high (HF), mid (MF) and low (LF).

Table 1 List of phrasal lexical items used in the test based on head verb frequency band

Restricted collocations	Frequency band	Total occurrences (of head)	Frequency of (the exact)PLIs in BNC
do things by halves	HLF	559 596	11
make tracks	HLF	217 268	31
take a fancy to	HLF	179 220	31

give NP the creep	HLF	131 417	30
keep a straight face	HLF	50 092	35
let NP into a secret	HF	29 768	23
join/enter the fray	HF	17 331	52
drive NP to drive	HF	16 477	8
act the goat	HF	15 620	6
avoid NP like the plague	HF	11 750	30
wipe NP off the map	MF	2 367	5
tighten NP's belt	MF	1 548	23
seal NP's fate	MF	1 512	16
spare no expense	MF	1 023	7
scrape the bottom of the barrel	MF	865	5
worship the ground NP walks on	LF	0	5
wring NP's neck	LF	0	24
pluck/summon up courage	LF	0	65
goad/spur NP into action	LF	0	28
toe the company line	LF	0	1

Results and Discussion

Individuals who have more idiomatic responses will also have higher plausibility scores for other responses.

The results show that the semantically plausible means for all 20 students. The highest mean achieved was 2.7. The results presented in Table 3 show means and standard deviations for all 20 respondents whose test results were coded.

Table 2 Percentage of idiomatic responses, semantic plausibility mean and standard deviation of individuals taking the test

Respondent	Idiomatic (%)	Idiomatic (20total)	Semantically plausible Mean	SD
Student 1	25%	5	1.75	1.18
Student 2	10%	2	1.53	0.94
Student 3	0%	0	1.70	1.39
Student 4	5%	1	1.87	0.95
Student 5	55%	11	2.58	1.80
Student 6	0%	0	1.90	1.23
Student 7	5%	1	1.64	0.78
Student 8	25%	5	2.7	1.66
Student 9	20%	4	2.09	1.30
Student 10	15%	3	1.48	0.50
Student 11	10%	2	2.17	1.50
Student 12	35%	7	2.18	1.27
Student 13	15%	3	1.98	1.20
Student 14	25%	5	2.23	1.28
Student 15	10%	2	2.08	1.44
Student 16	5%	1	1.76	0.97
Student 17	5%	1	1.42	0.84
Student 18	10%	2	1.6	0.65
Student 19	5%	1	1.15	0.32
Student 20	5%	1	1.92	1.11

Percentages of 'correct' idiomatic scores were categorized as high (>50%), average (15-25%) or low (0-5%). Only one respondent had a 'high' score, with 55% 'correct'. This student's mean

plausibility score was 2.58. About eight respondents were categorized as having low scores, with the rest categorized as average. However, the semantic plausibility metric seems to reveal that the respondents have different mean scores. These results led to performance of a correlation analysis. The correlation analysis was performed in order to observe the relationship between the two variables, i.e. number of idiomatic responses and mean semantic plausibility of non-idiomatic responses. The Pearson correlation coefficients (r) have values from -1 to +1. The motivation was to describe the strength and direction of the relationship between the variables. The following tables show the descriptive statistics of mean and standard deviation for the responses of 20 respondents for the study, and the summary of correlation results.

Table 3 Means and standard deviations for responses of 20 respondents

	Respondents (N= 20)	
	Mean	Standard deviation
Idiomatic responses	2.85	2.70
Mean of non-idiomatic responses	1.89	.38

Table 4 Summary of correlation results for idiomatic responses and mean non-idiomatic responses for 20 respondents

N=20	
Pearson correlation of idiomatic responses and mean of non-idiomatic responses	r = .652
p	.0009

The results revealed that the relationship between the number of idiomatic responses and the mean plausibility of the non-idiomatic responses was strong, with a positive relationship observed between the variables, r = .652, and, N = 20, p= .0009. This suggests that individuals who have higher numbers of idiomatic responses will also have higher plausibility scores on their non-idiomatic responses. This corresponds to the direction of the predicted correlation.

A case study - samples of an individual's set of responses

In this section a student's answers are presented and discussed in detail. Anna's non-idiomatic responses were analyzed using the mean results. Anna's (not her real name) answers were examples of high idiomatic responses. These responses were analyzed and compared to each other. The following table shows Anna's responses for both the idiomatic and non-idiomatic answers.

Table 5 Student 5 (T5) –Anna's responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Non-idiomatic response (Mean)	Standard Deviation
1.avoid	avoids	√		
2.act	play		5.7	
3.give	took		1.0	
4.drive	drive	√		
5.join/enter	join	√		
6.toe	following		4.0	
7.galvanise/ goad/spur	put		2.4	
8.worship	own		1.0	
9.do	do	√		
10.let	let	√		
11.make	made	√		
12.pluck/	plucked	√		

summon				
13.seal	twisted		1.7	
14.take	take	√		
15.keep	give		1.3	
16.spare	made		1.3	
17.wring	break		4.8	
18.tighten	tighten	√		
19.wipe	wipe	√		
20.scrape	scraped	√		
Total		11	2.58	1.80

Anna's score for the idiomatic responses was 55% or equal to 11 out of 20 idiomatic responses, and was the highest score of all 20 respondents. Her mean score was 2.58 for the non-idiomatic responses. This was the second highest among all respondents and this has led to analyzing all her semantically plausible responses. The following list shows her semantic plausibility verb selections for filling the gaps.

Anna filled in the gaps of the cloze test with 9 non-idiomatic answers, listed below:

1. ACT the goat- PLAY
2. GIVE NP the creeps- TOOK
3. TOE the company line- FOLLOWING
4. GALVANISE/ GOAD/ SPUR – PUT
5. WORSHIP the ground NP walks on- OWN
6. SEAL NP's fate- TWISTED
7. KEEP a straight face- GIVE
8. SPARE no expense- MADE
9. WRING NP's neck –BREAK

The non-idiomatic verbs were checked using a frequency rank in Kilgarriff's lemmatized BNC frequency list. There were 6,318 words in the lemmatized frequency list including 1,281 verbs. The verbs been grouped into bands with ten verbs per band, and Anna's responses were ranked based on these bands. The bands were restricted to the first 12 bands on the presumption that the first 120 verbs (10%) are at the high end of the frequency spectrum. The following table shows the appearance of the verbs based on the bands. Their frequency and sort order in Kilgarriff's lemmatized list are also listed in Table 6. An example of a verb band is given in Table 7.

Table 6 Band, sort order and frequency of the plausible verb responses

Verb	Band/12	Sort order	Frequency
PLAY	5	245	38,058
TOOK	1	54	179,220
FOLLOWING	4	203	46,145
PUT	2	125	69,978
OWN	>12	1,536	6236
TWISTED	>12	3,480	2004
GIVE	2	76	131,417
MADE	1	46	217,268
BREAK	11	532	19,512

Table 7 Band 1 of the first 10 verbs extracted from Kilgarriff's lemmatized BNC frequency list

N	Verb	Rank no
1	be	2
2	have	8
3	do	18
4	say	34
5	go	40
6	get	44
7	make	46
8	see	51
9	know	52
10	take	54

The results show that 7 plausible answers provided by Anna were among the 120 verbs in the first 12 bands. Only 2 verbs i.e. *own* and *twisted,* were below those bands in frequency. The results reveal that 77.8% of the plausible verb choices made by Anna 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 data presented in the first case supported Dornyei and Skehan's (2003) as well as Sawyer and Ranta's (2001) notion of the impact of individual difference on language learning in general. It will indirectly influence the vocabulary acquisition of each individual. There was evidence of a reasonably predictable knowledge of individual lexical items, shown in Table 2. What is more, the *idiomatic* column presented the respondents' knowledge of collocations and the mean semantic plausibility of non-idiomatic responses derived individually to reflect individuals' lexical knowledge. This observation is closely related to language processing. So, in this context, when the respondents were faced with a lexical retrieval task, they applied two possible strategies in retrieving specific vocabulary: either retrieving whole phrases or single words.

A non-native speaker like Anna does not have the advantages of native speakers whose number of fixed expressions stored in mental lexicon is vast (Jackendoff, 1995; Melčuk, 1995: 169), although it is hard to have an accurate estimation of the extent of the formulaic language stored in the mental lexicon (Kuiper, Columbus, & Schmitt, 2009). Kuiper, Columbus & Schmitt further argue that there is a possibility that there are a larger number of phrasal lexical items than single word items in a native speaker vocabulary. So, if learners like Anna, do not have sufficient PLIs in their mental lexicon, native-like competency is hindered, requiring her to opt for other strategies when having language difficulties.

The findings are significant because they illustrate the types of responses learners tend to come up with and indirectly illustrate the challenge of mastering restricted collocations. The study by Millar (2011) has provided support for the theory that malformed L2 collocations lead to an increased processing burden for native speakers in terms of slower reading speed. So, it does put some strain on native-speakers' processing. However, in the case of L2 and if we view L2 use from a lingua franca perspective, native-like attainment and selection may possibly not be the aim for L2 development. In this sense, the malformed or infelicitous restricted collocational choices made by L2 learners should be viewed more positively as instances of risk-taking strategy in order to cope communicatively.

References

Aisenstadt, E. (1981). Restricted collocations in English lexicology and lexicography *ITL Review of Applied Linguistics*, 53, 53-61.

Bahns, J. (1993). Lexical collocations: A contrastive view. *ELTJ, 47*(1), 56-63.

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

Biskup, D. (1992). L1 influence on learners' rendering of English collocations: A Polish/German empirical study. In A.A. Bejoint (Ed.), *Vocabulary and Applied Linguistics*. London: MacMillan.

Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2014). Gauging the effects of Exercises on verb-noun collocations. *Language Teaching Research*. *18* (1), 54-74.

Cowie, A.P. (1991). Multiword units in Newspaper Language. In S. Granger (Ed.), *Perspective on the English lexicon: A tribute to Jacques van Roey* (pp. 101-116). Louvain-la-Neuve: Cahiers de I'Institut de Linguistique de Louvine.

Cowie, A.P. (1992). Multiwords lexical units and communicative language teaching. In P. J. L.

Arnaud and H. Béjoint (eds), *Vocabulary and Applied Linguistics*. Houndmills: Macmillan, pp.1-12

Cowie, A.P. (1998). Introduction. In A.P. Cowie (Ed.), *Phraseology: theory, analysis and applications* (pp. 1-20). Oxford: Clarendon Press Oxford.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In *the Handbook of Second Language Acquisition*, C.J. Doughty and M.H. Long (eds), 589 630. Oxford: Blackwell.

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.

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*. Tubingen: Max Niemeyer Verlag.

Jackendoff, R. (1995). The boundaries of the lexicon. Idioms, structural and psychological perspective. In M. Everaert, E. van der Linden, A. Schenk, & R. Schreuder (Eds.), *Idioms: Structural and psychological perspectives*. (pp.133-166) Hillsdale, NJ: Erlbaum.

Kamariah Yunus, & Su'ad Awab. (2011). Collocational competence among Malaysian Law undergraduate students. *Malaysian Journal of ELT Research*, 7(1), 151-202.

Koya, T. (2004). Collocation research based on corpora collected from secondary school textbooks in Japan and in the UK. *Dialogue, TALK, 3*,7-18.

Koya, T. (2005). *The Acquisition of Basic Collocations by Japanese Learners of English.* Unpublished doctoral dissertation. Waseda University.

Koya, T. (2006). What is the reality of collocation use by native speaker of English? *Dialogue, TALK, 5*, 1-18.

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. (2009). *Formulaic Genres*. London: Palgrave-Macmillan.

Kuiper, K., Columbus, G., & Schmitt,N. (2009). Acquiring phrasal vocabulary. In S. Foster Cohen (Ed.), *Advances in Language Acquisition*. Basingstoke: Palgrave Macmillan.

Laufer, B., & Waldan, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning, 61*, 647-672.

Mel'čuk, I. (1995). Phrasemes in language and phraseology in linguistics. In M. Everaert, E. van der Linden, A. Schenk, & R. Schreuder (eds.), *Idioms: Structural and psychological perspectives*. (pp. 167-232) Hillsdale, NJ: Erlbaum.

Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics,* 32, 129-148.

Miyakoshi, T. (2009). *Investigating ESL Learners' Lexical Collocations; The Acquisition of Verb + Noun Collocations by Japanese Learners of English.* Unpublished doctoral dissertation. Hawai'i University, USA.

Moon, R. (1998). *Fixed Expressions and Idioms in English: A Corpus Based Approach*. Oxford: Clarendon Press.

Nesselhauf, N. (2003). Transfer at the locutional level: An Investigation of German-speaking and French–speaking learners of English. In Tschichold (ed.), *English Core Linguistics: Essays in honour of D.J. Allerton* (pp. 269-286). Bern: Peter Lang.

Nesselhauf, N. (2005). *Collocations in a Learner Corpus.* Amsterdam: John Benjamins.

Rafidah Kamarudin. (2013). *A Study on the Use of Phrasal Verbs by Malaysian Learners of English*. Unpublished doctoral dissertation. University of Birmingham.

Sawyer, M., & Ranta, L. 2001. Aptitude, individual differences, and instructional design. In *Cognition and Second Language Instruction*, P. Robinson (9ed.), 319-353. Cambridge: CUP.

Schmitt, N. Dörnyei, Z., Adolphs, S., & Durow, V. (2004) Knowledge and acquisition of formulaic sequences. In N. Schmitt (ed.), *Formulaic Sequences* (pp. 127-51). Amsterdam: John Benjamins.

Siyanova, A., & Schmitt, N. (2007). Native and non-native use of multiword Versus one-word verbs. *International Review of Applied Linguistics in Language Teaching 45*(2): 119 139.

Wray, A. (2002), *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, *44*, 647 668.