

Use of Phrasal Verbs in an ESL Learner Corpus and its Corresponding Pedagogic Corpus

Abdolvahed Zarifi^{1*} and Jayakaran Mukundan²

¹English Language Department, Yasouj University, Faculty of Humanities, Yasouj 75918-74934, Iran

²Department of Language and Humanities Education, Faculty of Language Studies, Universiti Putra Malaysia, Serdang 43400, Selangor, Malaysia

ABSTRACT

Phrasal verbs are one of the most notoriously puzzling aspects of English language instruction. Despite their potential complexities, they are of high relevance for ESL/EFL learners because knowledge of them is often equated with language proficiency and fluency. With the emergence of corpus linguistics, phrasal verbs have been extensively studied in General, Learner and Pedagogic corpora. Literature, however, is lacking in how learners' use of phrasal verbs reflects the corresponding pedagogic corpora to which they are exposed. To fill this research gap, this study adopted a corpus-based content analysis as its methodological approach to investigate the treatment of phrasal verbs in an ESL learner corpus and its corresponding pedagogic corpus. Findings are also compared against the presentation of these combinations in the British National Corpus (BNC). The study reveals that the selection of teaching materials is more intuitively than empirically based. It also suggests that teachers can use available corpora as supplementary teaching sources to work out the areas of L2 that tend to cause problems for the learners.

Keywords: Corpus linguistics, ESL context, learner corpus, pedagogic corpus, phrasal verbs

ARTICLE INFO

Article history:

Received: 18 October 2018

Accepted: 16 May 2019

Published: 18 December 2019

E-mail addresses:

vzarifi@yu.ac.ir (Abdolvahed Zarifi)

jayakaranmukundan@yahoo.com (Jayakaran Mukundan)

* Corresponding author

INTRODUCTION

With a number of studies in second language acquisition highlighting the significance of vocabulary and multiword expressions in language learning, the direction of a considerable body of linguistic research began to shift from syntax and phonology to the lexicon and multi-word expressions which are often looked at as the Cinderella

of the field (Zarifi, 2013). This paradigm shift was emphasized by Laufer (1997) observing, “After decades of neglect, lexicon is now recognized as central to language acquisition process”. Phrasal verbs, albeit viewed as the multiword middle ground between syntax and lexicon, are generally recognized as word-level entities and, thus, have come to the fore in both language studies and language instruction. The emergence of large electronic corpora and the development of robust software have enabled researchers to better identify these otherwise elusive structures (Hunston, 2002; Read, 2004). Phrasal verbs have been extensively studied in different types of corpora, including general corpora (Biber et al., 1999; Gardner & Davies, 2007), learner corpora (Schneider, 2004), pedagogic corpora (Akbari, 2009; Von, 2007; Zarifi, 2013; Zarifi & Mukundan, 2014a) and specialized corpora (Trebits, 2009). More specifically, in ESP field, comparisons are sometimes made between the use of phrasal verbs in corpora and sub-corpora to distinguish between different genres (Trebits, 2009). Likewise, native speakers’ use of phrasal verbs is often compared with that of the non-natives to identify the divide between the authentic language and the language that the textbooks try to model. However, to the best of the researchers’ knowledge, no attempt has ever been made to study how learner use of phrasal verbs is in keeping with the content of the pedagogic corpus, mainly the textbook materials that learners are exposed to as the main language input.

Study of Phrasal Verbs in Learner Corpora

There are a few corpus-based studies dealing with the use of phrasal verbs in learner and pedagogic corpora. For instance, Schneider (2004) compared the use of phrasal verbs in four sub-corpora of the International Corpus of English (ICE) from Singapore, the Philippines, India, and East Africa and compared them against the British English ICE corpus. The study was aimed at exploring how the speakers of these language varieties differed from one another in terms of the occurrence, frequency, structural behavior and productivity of these combinations. The findings indicated that the Singaporean learners enjoyed overusing phrasal verbs and tended to employ them considerably more than all the other varieties including the British English. Speakers of the other varieties, however, tended to underuse these combinations than native speakers. In addition, there appeared to be an inverse relationship between the level of stylistic formality and the propensity of the varieties to use phrasal verbs. To put this into perspective, while the use of phrasal verbs strongly featured spoken English in British and Singaporean English, their presentation in the other varieties turned out to be stylistically associated with more formal registers. In a similar way, Singaporean ESL learners tended to associate a wider range of potential word meanings with each phrasal verb unit than any other variety including BrE. While Indian and BrE English appeared to use each phrasal verb for an almost equal number of

meanings, learners from the Philippines and East Africa tended to associate phrasal verbs with a smaller number of word meanings.

In a similar way, Von (2007) investigated the use of phrasal verbs in the German and Italian components of the International Corpus of Learner English (ICLE) and compared the findings with Louvain Corpus of Native English Essays (LOCNESS). Findings revealed that German learners made use of more phrasal verbs than both the Italian and the native students. He reasoned that the German learners' overuse of phrasal verbs was motivated by their tendency to use more German verbs. Italian students, on the other hand, proved to underuse phrasal verbs simply because of their preference for Latinate verbs. His argument seems to be consistent with the fact that phrasal verb combinations are largely based on Germanic verbs and that these structures are absent from the Italian language.

In addition, in a corpus-based study, Akbari (2009) explored the Malaysian ESL learners' use of phrasal verbs in narrative compositions. The corpus of the study included a number of the Malaysian ESL students' narrative compositions sampled from the English of Malaysian School Students corpus (EMAS). Findings of the study revealed that the learners tended to avoid using the phrasal verb combinations of idiomatic type. In addition, they often experienced difficulty using these structures both in syntactic and semantic terms.

Moreover, Chen (2013), using a corpus of learner English and four native corpora of two English varieties, explored Chinese university students' use of phrasal verbs in

comparison with their American and British counterparts in argumentative and academic writing genres. The results showed that first, American students used phrasal verbs differently from their British counterparts both in terms of the number and variety of the combinations that they used. Not only did American students use more phrasal verbs in their argumentative and academic writing, but they also made use of a greater variety of phrasal verb combinations. Second, American and British students tended to use more phrasal verbs in argumentative writing than in academic writing. Simply because of these remarkable differences between the two groups of native students, the researcher found it really difficult to conclude whether the EFL Chinese learners had tendency towards over- or underuse of phrasal verbs in writing. Although the Chinese learners appeared not to be much different from the British students in terms of the overall frequencies of phrasal verbs they used, there existed some degree of discrepancy between the Chinese and American students in this regard.

Finally, Zarifi and Mukundan (2014b), in a corpus-based study of the use of phrasal verbs in the EMAS corpus investigated the issue of creativity and unnaturalness in the use of phrasal verbs by ESL Malaysian learners. Findings revealed that Malaysian ESL learners showed a great tendency towards both making up new phrasal verbs and overuse of these combinations in their language production. This enabled the researchers to argue that in spite of both the theoretical and empirical evidence "attesting the underuse of phrasal verbs

by non-native speakers, the Malaysian ESL learning context seems to counteract the effects of the differences between L1 and L2” (Zarifi & Mukundan, 2014b). Yet, there are some complicated features to the English phrasal verbs that tend to be problematic for nonnative speakers to grasp. While some of the forms they created appeared to be based on the existing patterns of phrasal verbs developed by the native speakers, and hence indicative of creativity among the non-native learners, some other combinations were really deviations of the standard use of the language both lexically and semantically. The learners’ use of the non-idiomatic combinations appeared to be almost error free, e. g. PUSH OUT, WALK OUT, etc.; however, the learners’ attempt to formulate new idiomatic forms was largely prone to error, e.g. SAY OUT, USE OUT, and VOICE OUT. As a result, the researchers came to conclude that the English language tends to keep the mystery of creating new idiomatic phrasal verbs as a linguistic property to its own native speakers and hide it from the non-native speakers.

Study of Phrasal Verbs in Pedagogic Corpora

Because of the overwhelming number of phrasal verbs in English and the different collections available to choose from the right selection and presentation of these combinations have frustrated curriculum designers and materials developers. This sense of frustration still gets more complicated by the fact that course books cannot in any possible way include all

these expressions and neither are all of these combinations equally useful to the learners. Therefore, of main research interest is whether curriculum developers are really taking into account the corpus-based research findings in selecting and presenting these structures or are only paying lip service to the use of empirical findings in teaching materials.

In keeping with the above research concern, the selection and presentation of the phrasal verb combinations in the ELT materials has been investigated in different pedagogic corpora. Examining the presentation of phrasal verbs in a few ELT course books and reference materials, Side (1990) contended that some of the difficulties that learners encountered in dealing with these combinations were motivated by the way in which they were presented. She argued that ELT materials often failed “to create learnable patterns” of the combinations and they sometimes present them in “patterns of the wrong kind”. While it is the particle element that gives some lexical verbs a specific aspect of meaning, course books often formulate lists that revolve around a particular lexical verb. On the other hand, dictionaries sometimes assign meanings to the particle elements of the phrasal verbs that are more conveyed by the lexical verbs. For instance, Longman English Grammar defines ‘UP’ as “confining/fastening/mending, etc.” exemplified by ‘LOCK STH UP, STICK STH UP, PACK STH UP, etc.’. Much of this sense is, however, conveyed by the lexical verb rather than the particle.

In addition, Darwin and Gray (1999) developed a list of the 20 most frequently occurring phrasal verbs in the BNC. Comparing the list with the phrasal verbs in a typical ESL grammar book, they found that only 3 of all the phrases in the textbooks matched the 20 phrasal verbs on the list. Likewise, Koprowski (2005) studied the treatment of phrasal verbs and other multiword expressions in three contemporary ELT course books. Although they were all developed as British general English materials for learners at the intermediate level, not even a single phrasal verb was found to be shared by the three textbooks. In addition, less than one per cent of the multiword expressions were shared by any of the course books. Lamenting the lack of consistency among the textbooks, Koprowski observed that ELT materials developers did not follow any principled criteria in the process of vocabulary selection.

In a recent study, Zarifi and Mukundan (2015) investigated the semantic treatment of these phrasal verbs in a corpus of ESL textbooks. Using WordSmith software and the Oxford Dictionary of Phrasal Verbs as research instruments, the study revealed that most of these combinations were presented with a very thin skeleton. Despite the huge number of phrasal verbs in the corpus, they turned out to be inadequately dealt with in terms of quality. Some combinations were presented with their rare and highly infrequent word meanings. Many others were narrowly presented in different Forms with the same meaning.

In another recent study on the Malaysian ESL textbooks, Zarifi and Mukundan (2012) investigated the use of phrasal verb combinations in the spoken sections of the materials. They found that both the selection and presentation of these structures were inconsistent with their natural use in the BNC. They reported that textbooks did a disservice to the learners by including combinations of extremely low frequency counts in general English and excluding a number of combinations with highly low frequency counts in general English.

The corpus-based study of the presentation of phrasal verbs in different pedagogic corpora has in fact provided illuminating findings on different aspects of these structures in different instructional language varieties. In a similar way, results of the empirical studies of ESL learners' use of phrasal verbs show how different ESL learners tend to use them, and how their treatment of these forms is in dis/agreement with their use by native speakers. The literature is, however, lacking in any cross-corporal study comparing how learner use of phrasal verbs mirrors the presentation of these forms in the corresponding pedagogic corpus they get exposed to as their main language input. In other words, there exists no evidence as to whether ESL/EFL textbook content is in keeping with learner needs. The current study was, therefore, carried out to bridge this research gap in the literature. More specifically, this cross-corporal study aimed to address the following questions:

1. What are the phrasal verbs in each of the ESL Learner and Pedagogic corpora?
2. How does the use of phrasal verbs differ in the two corpora?
3. To what extent is the use of phrasal verbs in the two corpora consistent with their presentation in the BNC?

METHODS

This study is a corpus-based content analysis of the use of phrasal verbs in two ESL corpora namely, the EMAS and Malaysian ESL secondary level textbooks. The EMAS is a learner corpus of Malaysian ESL School Students created by Universiti Putra Malaysia. It contains around 472,652 tokens and includes written essays of 872 students at the secondary level and primary level. In order for the corpus to represent the southern- central, the northern, and the eastern parts of peninsular Malaysia, the respondents were selected from the schools in three states. These schools had performed well in the 1999 standardized examinations. This sample was intentionally selected in order to obtain sufficient and appropriate data for analysis. In other words, students with low proficiency in the English language were excluded as they were not expected to provide enough language data for the establishment of a corpus needed for further analysis. The data in this learner corpus were collected in the form of three narrative essays which each learner was required to write. The first essay dealt with a series of pictures, depicting a number of events happening to some kids going on a picnic

to the river bank, of which the students were required to provide a written account within one hour. For the second essay, the students were asked to write an essay on the topic 'The happiest day of my life'. The third essay involved an essay selected from the essays that the teachers had assigned to the learners as part of their regular school work. The selection of these essays was motivated by their potentiality to generate extensive language, familiarity to the students, as well as control of language support. It should, however, be pointed out that for comparison purposes the current study does not include the data collected from the primary level students. In other words, the learner corpus consisted of only 370,876 running words, extracted from a total number of 1507 narrative essays that were produced by 567 students from the secondary level.

The pedagogic corpus, on the other hand, includes the five textbooks of the English textbooks prescribed for the Malaysian students at the secondary level. It contains 302,642 tokens and comprises an almost balanced selection of texts in terms of spoken versus written modalities and a variety of general topics. Despite the use of phrasal verbs in different parts of each lesson in the whole corpus, a few lessons particularly deal with introducing and teaching phrasal verbs. The selection of these textbooks as the pedagogic corpus in this study was, in fact, informed by the observation that textbooks form "the core of most teaching programs" (Brown, 2011), and empower the learners to consolidate their language learning both inside and outside the classroom (Mukundan, 2004).

It should be pointed out that the selection and presentation of phrasal verbs in this pedagogic corpus appear not to be informed by the findings of corpus-based studies. For instance while the Form One textbook fails to include such highly frequent forms as ‘BREAK OFF, HOLD UP, COME THROUGH’, etc., it contains some rare items like ‘LAZE AROUND, POKE ABOUT, WHAM BACK’, and so on. Moreover, the textbook fails to make a distinction between Verb + Particle and Verb + Preposition structures, presenting phrasal verbs like PUT ON with prepositional verbs like ‘think about’, ‘concentrate on’, etc., (Form Four) and even with prepositional constructions like ‘responsible for’ and ‘happy with’ (Form Five). This might mislead the students to conclude that structures that come under the same category and appear together in the same lesson behave similarly. However, while ‘ON’ in ‘CONCENTRATE ON’ is a preposition, forming a semantic unit with its following NP, ‘ON’ in ‘PUT ON’ is a real particle, making a semantic whole with the verb not with its following NP.

For data collection, the WordList function of the WordSmith software version 5.0 (Scott, 2008) was used to extract all the potential particle elements in the corpora. Then the Concordance function was run to locate all the instances of particle elements preceded by an adjacent or a nonadjacent lexical verb. It can be said that the study adopted a comprehensive sampling as all the phrasal verb combinations in different sections of the textbooks were included. Since a verb particle might behave

differently as an adverb, a preposition, or a verb., based on the context in which it appears, we settled for a clear-cut definition that would differentiate phrasal verbs from prepositional verbs and other similar combinations. This study defines a phrasal verb “as combination of a lexical verb and a non-prepositional particle element that is either adjacent or nonadjacent to the verb” (Zarifi, 2013).

Following the definition, all the instances of lexical verbs followed by a particle of non-prepositional meaning (e.g. TURN ON) were tagged as phrasal verbs (VPart) to be distinguished from prepositional verbs (e.g. LOOK AT). Then, all the extracted combinations were lemmatized (turn, turned, turning = TURN) to have all the inflectional forms of each phrasal verb counted together. In order for the tagging process to be valid and acceptable, we asked an independent coder to read through the concordance lines and code all the combinations. The second coder who held a PhD degree in TESL had a record of teaching English as a second language for the past five years. Finally, Cohen’s Kappa statistics was used to check for the consistency level of coding process between the coders. The inter-rater reliability value was calculated to be 0.94, which is an excellent index of coding consistency.

In the last stage, we compared all the combinations extracted from the two corpora. In a similar way, the findings were compared against the BNC as the reference corpora to see how the two ESL corpora agreed with actual language use of these combinations.

RESULTS AND DISCUSSION

Application of the WordList and Concord functions of the WordSmith software (Scott, 2008) yielded the following results:

Table 1
General descriptive statistics of phrasal verbs in the Malaysian ESL pedagogic corpus

PV tokens	PV Lemmas	LV Lemmas in PVs	Particle elements in PVs
2212	464	226	23

As Table 1 reveals, a combination of 226 different lexical verbs with 23 different particles made up a total number of 2212 tokens of phrasal verbs in the Malaysian pedagogic corpus. These combinations appeared in 464 different lemmas. In other words, on the average, each lemma appeared roughly in 4.77 tokens of phrasal verbs, and each lexical verb took part in the formation of 9.79 phrasal verbs.

Table 2
General descriptive statistics of phrasal verbs in the EMAS

PV tokens	PV Lemmas	LV Lemmas in PVs	Particle elements in PVs
2120	492	244	23

Table 2 shows that a combination of 244 different lexical verbs with 23 different particle elements resulted in a total number

of 2120 tokens of phrasal verbs in the Malaysian ESL learner corpus. These combinations appeared in 492 different phrasal verb lemmas. On average, each phrasal verb lemma provided 4.31 tokens of phrasal verbs and each lexical verb accounted for the formation of 8.68 tokens of phrasal verbs. It should be pointed out that the learners' incorrect use of phrasal verbs was also included for analysis. For example, nonstandard use of PICK UP in the utterance 'PICKED UP the flowers' and many other categorically ill-formed combinations like STATE OUT were counted as phrasal verbs.

In order for the figures to be directly comparable, they were normalized. To this end, the observed frequency counts of phrasal verbs in each corpus were projected to a corpus basis of exactly 10,000 words. First, Chi-square statistics revealed that the two corpora did not significantly differ in terms of the number of phrasal verb tokens they contained at 0.01 level of significance. Second, the normalized figures in Table 3 show only some negligible degree of difference in distribution of phrasal verb lemmas and lexical verb lemmas between the two corpora. That is, while in the pedagogy corpus, 15 phrasal verb lemmas and 8 lexical verb lemmas appeared per 10,000 tokens, about 13 phrasal verb lemmas and 7 lexical verb lemmas appeared per 10,000 tokens in the EMAS.

Table 3
Frequencies normalized to a 10,000-token corpus

Corpus	Token	PV (Norm)	PV Lemmas (Norm)	LV Lemmas (Norm)
EMAS	319,725	2120 (66)	492 (15)	244 (8)
Pedag.	302,642	2212 (73)	464 (15)	226 (7)

Care should, however, be exercised in the interpretation of the noticeable similarity between the two corpora. While some units (23 forms) of common use in textbooks like FILL IN, FIND OUT, and SOUND OUT, (Figure 1) account for the extensive use of phrasal verbs in the pedagogic corpus, the overuse of these combinations in the learner corpus should be interpreted in terms of the fact that the Malay language, unlike English, is a verb-framed rather than a satellite-framed language (Talmy, 1991). In other words, while the direction of the motion in English is conveyed through particle elements (satellites) like into, down, out, in, up, and down, in the Malay language this notion is expressed by the verb such bertembung= run into, jatuh= fall down, keluar= go out, masuk= go in, and naik= go up, turun =go down. By the same token, we tend to expect ESL learners with verb-framed language backgrounds like Malay to avoid using phrasal verbs (Liao & Fukuya, 2004) which are common in satellite-framed languages like English.

Further data analysis revealed that a number of items like TAKE OVER, WORK OUT, BRING ABOUT, BRING IN,

BRING DOWN, GET IN, COME ABOUT, TURN OFF, MOVE BACK, PICK OUT, etc., that appeared in the pedagogic corpus were absent from the learner corpus. The absence of these items in the learner corpus, however, needs to be dealt with cautiously. To begin with, some of these items like TAKE OVER, WORK OUT, and BRING ABOUT, are highly specialized and enjoy a limited meaning coverage. To follow, the absence of these units in the learner corpus can be partly accounted for in terms of the number of the topics it dealt with. Unlike the pedagogic corpus that dealt with a large number of topics, the learner corpus was developed by the respondents writing about only three topics. Finally, although these combinations occurred in the pedagogy, they were, nonetheless, not repeated enough and recycled appropriately to be imprinted in the learners' mind. They occurred for less than three times and were not recycled at suitable spaced intervals to consolidate learning and motivate productive use. In other words, these items were absent from the learner corpus either because the learners did not find the chance to use them, or because they failed to learn them well due to the

N	Concordance	Set	Tag	Word #	t. #	os.	#	os.	#	os.
1	Strengthen Children Self.html and find out the definition of self-esteem and			2,082	165	9%		0	8%	
2	weight? TASK 2 Try this simple quiz to find out whether you are at risk of having			669	60	5%		0	2%	
3	My group interviewed 50 ex-smokers to find out why they had stopped smoking.			3,917	267	3%		0	0%	
4	this issue. Interview smokers to find out why they continue to smoke.			3,268	209	5%		0	7%	
5	INFO BYTE Log on to the Internet to find out what measures other countries			1,727	120	3%		0	5%	
6	environment. Do an Internet search and find out the following: • things that are			3,711	238	1%		1	0%	
7	• Value for money Example: You may find out that there is no price list for food			3,633	234	7%		1	8%	
8	Click on the following websites and find out how they can help consumers.			3,585	233	7%		1	7%	
9	should start from your home or school. Find out the aspects where consumer			3,523	229	0%		1	6%	
10	in the sale. It is important that you find out what you are getting for your			1,278	97	7%		1	2%	

Figure 1. A concordance snapshot of the use of 'find out' in the pedagogic corpus

manner of presentation and practice of the items by the authors or by the teachers in the classroom. It seems also possible that they managed to choose another non-phrasal verb form simply because they felt that it would fit the context better than the phrasal verb form.

In order to have a clearer picture of the way phrasal verbs were presented in the two corpora, the researchers decided to compare the top twenty phrasal verbs in them and also against the top 100 phrasal verbs in

the BNC. Table 4 shows the top 20 phrasal verb lemmas in the BNC, EMAS and the Pedagogic corpora.

As Table 4 shows, while 4 of the top 20 phrasal verbs in the BNC, i.e., POINT OUT, COME OUT, COME IN, and GET BACK, did not appear in the pedagogic corpus, only 3 of the top 20 items in the BNC, i.e., WORK OUT, TAKE OVER, POINT OUT, were absent from the learner corpus. In a similar way, while 9 of the top 20 phrasal verbs in the pedagogic corpus

Table 4
Top 20 phrasal verbs in the BNC, Malaysian ESL pedagogy and EMAS

Phrasal Verb Lemma	BNC Rank	Pedagogy Rank	Learner Rank
Go on	1	28	12
Carry out	2	3	#
Set up	3	18	10
Pick up	4	10	4
Go back	5	25	2
Come back	6	61	6
Go out	7	12	5
Point out	8	*	*
Find out	9	1	14
Come up	10	31	97
Make up	11	32	99
Take over	12	26	*
Come out	13	*	7
Come on	14	37	46
Come in	15	*	35
Go down	16	76	28
Work out	17	4	*
Set out	18	79	#
Take up	19	20	#
Get back	20	*	20
Fill in	*	2	#
Write out	*	4	#
Look up	26	5	#

Table 4 (continue)

Phrasal Verb Lemma	BNC Rank	Pedagogy Rank	Learner Rank
Look after	*	6	#
Pick out	75	7	#
Write down	*	8	#
Take part	*	9	#
Cut down	*	11	#
put up	33	13	#
Wake up	*	14	1
Go through	73	15	#
Throw away	*	16	#
Give up	24	17	#
Check out	*	19	#
Fall down	*	#	3
Pull up	*	#	8
Get up	25	#	9
Bring along	*	#	11
Turn back	56	#	13
Calm down	*	#	15
Bring up	36	#	16
Pass away	*	#	17
Grow up	*	#	18
Put down	32	#	19
Bring back	40	#	20

* zero frequency count in the corpus

present in the corpus but not included within the range of the rank specified

did not appear in the top 100 phrasal verbs in the BNC, only 7 of the top 20 phrasal verb combinations in the learner corpus were absent from the top 100 units in the BNC. In addition, COME OUT and GET BACK, which were totally absent from the pedagogic corpus, were among the top 20 phrasal verbs in the BNC and the learner corpus. Although the disappearance of the high frequent items like COME OUT and GET BACK from the pedagogic corpus can be raised as a shortage of the textbooks, the absence of combinations like TAKE OVER and WORK OUT in the learner corpus could be attributed to the level of formality of these combinations.

Findings also showed that combinations like FILL IN, WRITE OUT, LOOK AFTER, WRITE DOWN, TAKE PART, THROW AWAY and CHECK OUT were among the high frequent items in the pedagogic corpus but not so in the learner and the BNC corpora. While some of these items like FILL IN, WRITE OUT, WRITE DOWN, and CHECK OUT are the expressions specific to pedagogic register, the overuse of the other forms indicates that the development of the textbook materials failed to mirror the natural language use. In addition to the absence of some top frequent phrasal verbs from the pedagogic corpus, the corpus included some highly infrequent forms like 'WHAM BACK, RUSTLE OUT, and SPROUT OUT' for which BNC query cropped out zero or some negligible frequency shots. On the other hand, there were some top frequent PVs in the learner corpus like FALL DOWN and PASS AWAY

that were not so in the Pedagogic corpus and the BNC. This discrepancy can be interpreted in terms of the size and makeup of the Pedagogic corpus and the nature of the topics it dealt with. For instance, one section of the corpus required the respondents to describe a series of pictures in which somebody falls into the water, hence overuse of FALL DOWN (Figure 2). A wider range of topics would yield a different and perhaps better picture of the learners' use of the PVs and, by implication, other forms of the language.

Moreover, a quick look at the phrasal verbs in Table 4 shows that the items most used by the learners are those that have more transparent meanings in their lexical and/or particle elements (e.g., 'go back', 'wake up', and 'fall down'), and the more non-compositional items (e.g., 'carry out', 'take part', and 'take over') are lower in frequency or even absent from the learner corpus. Among others, one key implication of this finding is that learners do show evidence that they somehow acquire phrasal verbs in spite of their being absent from the pedagogic corpus. This seems to be especially the case with the phrasal verbs that enjoy more transparent rather than opaque semantic properties. Therefore, with pedagogical issues in mind, perhaps one important takeaway could be that course-books writers should not heed the BNC and BNC-like corpus data willy-nilly. In other words, learners tend especially to benefit from overt attention being drawn to these more non-compositional phrasal verbs since the evidence suggests that the learners may not simply pick them up as readily.

	Set	Tag	Word #	t. #	os.	#	os.	#	o: ^
1			so surprise and keep seeing to book to fall down from the shelves, then, the last	84	8 4%		0 1%		
2			instruction. But I was so careless and fall down from the tree. My leg was	310	26 7%		0 8%		
3			girls were very careless. The girl who fall down into the river do not know how	64	5 3%		0 6%		
4			some flower. Suddenly, one of the girl fall down into the river. Both of the girls	49	3 4%		0 5%		
5			She continues her job. Suddenly, she fall down to the river. When Mary see	121	13 0%		0 8%		
6			our journey. "Be careful, Ling Ling don't fall down" I heard Pei Yin shouted.	136	14 4%		0 8%		
7			the river because the river bank was fall down. Syafinaz was panic and don't	116	8 6%		0 6%		
8			hear my sound. Suddenly, Roza was fall down in the alke. Ive geat a shock	137	9 6%		0 5%		
9			meet that two girls. Abrupt, Sheila fall down into the lake. She shout for	160	16 0%		0 0%		
10			back and saw one of the girls was fall down in that river. They quickly get to	101	7 5%		0 6%		

Figure 2. A concordance snapshot of the use of 'fall down' in the EMAS corpus

Of major pedagogical concern to ESL materials developers should be the combinations like HOLD BACK, GIVE IN, and BREAK OFF, which were used by the ESL learners and are included among the top 100 forms in the BNC but were, nevertheless, missing in the pedagogic corpus. On the one hand, just because these items do not appear in the pedagogic corpus, chances are that very little or no formal teaching of these combinations occurs in the classroom. On the other hand, the appearance of these items in the learner corpus shows that since English serves as a second language in Malaysia, language learners tend to pick up part of their language knowledge through encounter with sources other than textbooks or outside the classroom context. Given sufficient out-of-class input, it can be argued that even coursebooks cannot foil the Zipfian frequency distributions that students will engage with. Similarly, the data show that teachers should not be so much concerned with the (relatively limited) language input presented in the prescribed textbooks, and instead should encourage as much meaningful engagement with the

L2 as possible, since there exists evidence that students are also sensitive to naturally occurring frequency distributions, hence mitigating the concern with course book syllabi.

Of equal interest to materials developers should be the top frequent items and rare forms like WHAM BACK, RUSTLE OUT, and SPROUT OUT, in the pedagogic corpus that are either infrequent or missing in the learner and the BNC corpora. This shows that if authors rely on their intuition for the selection of teaching materials, they might fail to provide materials that are really appropriate to the target learners. They might either provide materials that are rare in the language or avoid materials that can be really useful to the learners. All these empirical findings have led the researchers to suggest that corpus-informed materials are more likely to take care of the learners' needs and preferences. Likewise, the formal instruction, if directed towards the high frequent forms and structures, can accelerate the acquisition of these forms in particular, and foster the process of language learning in general.

Findings also showed that a large number of top frequent items like HOLD UP, HOLD OUT, COME ROUND, SET ABOUT, COME OFF, SET DOWN, and MOVE UP, in the BNC were absent from both the pedagogic and learner corpora. While it seems insensible to question the learners for not using these frequent items, the absence of these combinations in the pedagogy corpus is a different story. In other words, we would like to concur with rationalists and argue that the absence of an element in a learner corpus does not prove the nonexistence of that item in the internalized system of the learner. The exclusion of these combinations from the pedagogy can, however, be thrown into question on the grounds that “dramatic differences in frequency should be among the most important factors influencing pedagogical decisions” (Biber & Conrad, 2001).

Finally, unlike the phrasal verbs in the pedagogic corpus, the combinations in the learner corpus seemed to follow closely the actual use of these structures in the BNC. To put this into perspective, while only 12 of the top 20 phrasal verbs in the pedagogic corpus were included in the top 100 phrasal verbs in the BNC, 14 instances of the top 20 items in the learner corpus were covered in the top 100 phrasal verbs in the BNC. More interestingly, PICK UP, COME BACK and GET BACK, which respectively ranked 4th, 6th and 20th in the BNC, occurred exactly with the same rank in the learner corpus. In the same way, GO ON, SET UP, GO BACK, GO OUT, FIND OUT and COME

OUT which come the 1st, 3rd, 5th, 7th, 9th and 13th in the BNC ranked the 12th, 10th, 2nd, 5th, 14th and 7th in the learner corpus. On the other hand, the closest rank pattern of phrasal verbs shared by the BNC and the Pedagogic corpus appeared to belong to CARRY OUT, TAKE UP, PICK UP, GO OUT, FIND OUT and WORK OUT which ranked the 2nd, 19th, 4th, 7th, 9th and 17th in the BNC. These combinations came respectively in the 3rd, 20th, 10th, 12th, 1st and 4th in the pedagogic corpus.

CONCLUSIONS

Results of the study enabled the researchers to conclude that: first, ESL textbooks seem to be developed on an ad hoc basis, more based on the writers’ intuition and speculation than on empirical findings; second, there is a divide between the language ESL learners pick up and the textbook language they get exposed to; third, learners in ESL contexts tend to incidentally pick up language through exposure to sources other than the textbook language; and finally, in order to have a better corpus of real life production of the learners of the language, learner corpora compilers should have the respondents write on a variety topics of their own interest rather than confining them to writing on specific topics. This would help them to produce more creatively and freely and cause their production to better represent their active knowledge of the language.

It is hoped that freely available web-based corpora like the BNC, the LOCNESS, and the Corpus of Contemporary American English (COCA), could be

used as supplementary teaching sources in classrooms, whereby language teachers can compare and locate the differences between learners' language and the target language and work out the areas of L2 that are likely to cause problems for the learners. In addition to their usefulness in error correction, corpora have a priceless pedagogic potential to, in Sinclair's words (Sinclair, 1997), "clarify, give priorities, reduce exceptions and liberate the creative spirit" of the language learners. More specifically, corpora can empower the teacher in grading and rearranging the course materials so that instructional materials include both the most frequent items and those tending to create difficulty for students. It is also hoped that materials designers take effective steps to develop new textbooks that comply with the natural use of language as evidenced in corpora. A case in point could be the inclusion and explicit teaching of the top 100 frequent phrasal verbs at least in the early stages of language teaching and learning. Because of their high frequency counts and, by the same token, for their usefulness to the learners, these items should be considered a priority in pedagogy and brought into sharp focus for explicit teaching and learning (Kennedy, 2002).

Despite its findings, this study suffers from some limitations. First, the learner corpus (EMAS) is based simply on three timed writing tasks for which the learners did not apparently have much of any choice in terms of topic. Second, although a pedagogic corpus consists of all the language data that

a learner gets exposed to in the classroom, the pedagogic corpus in this study included only the English textbooks prescribed for the Malaysian students at the secondary level. Yet, the inclusion of the textbooks seems to be reasonable as they serve as the main source of language materials in Malaysian ESL formal education (Nooreen & Arshad, 2005), and all the class activities are largely motivated by the contents of the textbooks.

This study is likely, we hope, to provide a strong incentive for further research in ESL teaching and learning. One area of interest could be evaluating various ESL teaching materials against the general corpora. Another area of main research interest could be comparing different ESL textbooks with their corresponding ESL learner corpora. Studies of such nature would be highly promising due to the pivotal role of instructional materials in ESL learning contexts. It is highly likely that the language input in textbooks accounts for some of the difficulties that ESL learners experience in picking up such challenging aspects of the language as phrasal verbs.

ACKNOWLEDGEMENT

The present paper has not been financially supported by any party, but we are grateful to the reviewers for their comments on an earlier version of the manuscript.

REFERENCES

- Akbari, O. (2009). *A corpus-based study on Malaysian ESL learners' use of phrasal verbs in narrative compositions* (Unpublished Doctoral thesis), Universiti Putra Malaysia, Malaysia.

- Biber, D., & Conrad, S. (2001). Quantitative corpus-based research: Much more than bean counting. *TESOL Quarterly*, 35(2), 331-336.
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English*. Cambridge, England: MIT Press.
- Brown, D. (2011). What aspects of vocabulary knowledge do textbooks give attention to? *Language Teaching Research*, 15(1), 83-97.
- Chen, M. (2013). A corpus study of English phrasal verb use by Chinese, British and American university students. *International Journal of Corpus Linguistics*, 18(3), 418-442.
- Darwin, C. M., & Gray, L. S. (1999). Going after the phrasal verb: An alternative approach to classification. *TESOL Quarterly*, 33(1), 65-83.
- Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *ELT Journal*, 35(3), 339-359.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge, England: Cambridge University Press.
- Kennedy, G. (2002). Variation in the distribution of modal verbs in the British National Corpus. In R. Reppen, S. Fitzmaurica & D. Biber (Eds.), *Using Corpora to Explore Linguistic Variation* (pp. 73-90). Amsterdam, Netherlands: John Benjamins.
- Koprowski, M. (2005). Investigating the usefulness of lexical phrases in contemporary coursebooks. *ELT Journal*, 59(4), 322-332.
- Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rational for pedagogy* (pp. 20-52). Cambridge, England: Cambridge University Press.
- Liao, Y., & Fukuya, Y. (2004). Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning*, 54(2), 193-226.
- Mukundan, J. (2004). *A composite framework for ESL textbook evaluation* (Doctoral thesis), Universiti Putra Malaysia, Malaysia.
- Nooreen, N., & Arshad, A. S. (2005). Examining the importance of EST and ESL textbooks and materials: Objectives, content and form. *ESP World: Online Journal for Teachers*, 9(4).
- Read, J. (2004). Research in teaching vocabulary. *ARAL*, 24, 146-161.
- Schneider, E. W. (2004). How to trace structural nativization: Particle verbs in world Englishes. *World Englishes*, 23(2), 227-249.
- Scott, M. (2008). *Wordsmith Tools (Version 5)* [Computer software]. Liverpool, England: Lexical Analysis Software Ltd.
- Side, R. (1990). Phrasal verbs: Sorting them out. *ELT Journal*, 44(2), 144-152.
- Sinclair, J. M. (1997). Corpus evidence in language description. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and language corpora* (pp. 27-39). New York, USA: Addison Wesley Longman.
- Talmy, L. (1991). Path to realization: A typology of event conflation. *Proceedings of the Seventeenth Annual Meeting of the Berkeley*, 17, 480-519. <http://dx.doi.org/10.3765/bls.v17i0.1620>
- Trebits, A. (2009). The most frequent phrasal verbs in English language EU documents - A corpus-based analysis and its implications. *System*, 37(3), 470-481.
- Von, V. (2007). *Phrasal verbs in learner English: A corpus-based study of German and Italian students* (Unpublished doctoral thesis), Albert-Ludwigs-Universität, Germany.

- Zarifi, A. (2013). *Establishing and evaluating phrasal verb use in a Malaysian secondary school textbook corpus* (Unpublished Doctoral thesis), Universiti Putra Malaysia, Malaysia.
- Zarifi, A, & Mukundan, J. (2012). Phrasal verbs in Malaysian ESL textbooks. *English Language Teaching*, 5(5), 9-18.
- Zarifi, A, & Mukundan, J. (2014a). An alternative approach to classification of phrasal verbs. *IUP Journal of English Studies*, 9(1), 114-127.
- Zarifi, A, & Mukundan, J. (2014b). Creativity and unnaturalness in the use of phrasal verbs in ESL learner language. *3L: Language, Linguistics, Literature®*, 20(30), 51-62.
- Zarifi, A, & Mukundan, J. (2015). A corpus-based study of semantic treatment of phrasal verbs in Malaysian ESL secondary school textbooks. *Pertanika Journal of Social Sciences & Humanities*, 23(4), 793-808.