Phrasal Verbs in English Textbooks Used in China's High School: A Multicorpus-Based Analysis

Wenting Yang

***Abstract*—Phrasal verbs (PV) are frequently used by native speakers of English. As a result, it is critical to include PVs in teaching routine to prepare students to be fluent English speakers. However, PVs are very difficult to teach due to their polysemic nature. In order to find out if China's textbooks contain enough frequently-used PVs, the essay compares PVs in China's textbooks with those in British National Corpus (BNC) and Corpus of Contemporary American English (COCA).**

***Index Terms*—Corpus, English, phrasal verbs, textbook.**

I. Introduction

The learning and teaching of phrasal verbs (PVs) are widely considered as one of the most challenging tasks of English Language Teaching (ELT) (Cornell, 1985; Kurtyka, 2001; Kartal, 2018) [1]-[3]. The difficulty lies in the nature of PVs: (1) polysemous; (2) complex grammatical construction and collocation; and (3) vary in meanings and forms as the language develops (Kurtyka 2001) [2]. Nevertheless, no matter how challenging it is, efforts should be made to delve in the approaches to tackle the problems as PVs are very common and productive in English language (Chen, 2007; Condon, 2008; Garnier & Schmitt, 2015; Kartal, 2018; Liao & Fukuya, 2004) [3]-[7]. However, in reality, non-native speakers tend to avoid using PVs when speaking English (Hulstijn & Marchena, 1989) [8]. Chen (2007) [4] provides several explanations, one of which is the lack of high-quality teaching and learning materials.

For China's high school students, their main English learning materials are their English textbooks. It is because in China's exam-oriented environment, students do not have much time to do extracurricular reading to acquire vocabularies incidentally. As a result, English textbook is the primary source for them to acquire English vocabularies. However, vocabularies presented in the textbook may not reflect the actual language usage in real life. According to Darwin and Gray (1999) [9], instructors, curriculum designers and researchers usually rely on their intuition, instead of lexical frequency, to determine the presentation of PVs. That is not scientific, because errors are likely to generate due to the arbitrary nature of human intuition. On the other hand, corpus-based analysis is recognized as being able to demonstrate a "detailed view of how real people speak and write in everyday situations" (McCarthy, 2004) [10], and thus has high potential to be used as a more reliable reference for textbook development.

The aim of the research is to draw a comparison between the PVs shown in China's PEP high-school English textbooks and Liu's (2011) [11] list of 150 PVs – 104 PVs from Biber, Johansson, Leech, Conrad & Finegan (1999) [12] and Gardner & Davies' (2007) [13] combined list, and 48 PVs from Liu's (2011) research. Moreover, Liu (2011) [11] combined two pairs of PVs to decrease the number from 152 to 150.

Additionally, it should be noticed that keeping consistency of the definition of PVs between Liu's list and the TC is significant. It is because there are too many theories proposed to define PVs (Liu, 2011) [11]. Different definitions are likely to produce different results. Since Liu (2011) [11] followed Gardner and Davies' (2007) [13] definition to generate the combined list, the research will also use the same definition to keep the consistency.

Research question:

(1) Are all the 150 frequently-used PVs in BNC and COCA included in the PEP textbook corpus (TC)?

(2) Do PVs in TC has similar rank order as those in BNC and COCA?

II. Motivation for Creating and Using this Type of Corpus

Garnier & Schmitt (2015) [6] summarized that PVs are important and difficult to learn for four reasons.

(1) PVs are very common in language use. Gardner & Davies (2007) [13], based on their findings in BNC, estimate that English learners will encounter one PV in 150 words.

(2) PVs are polysemic and functional. Gardner & Davies (2007) [13] found that each of the most frequent PVs had 5.6 meaning senses on average.

(3) Using PVs is significant to be fluent in English and to sounding as a native.

(4) PVs are composed of two or more orthographic words, thus English learners may decode the meanings of PVs' from their individual components instead of treating the PVs as single semantic units, which may lead to the misinterpretation of the PVs.

In a word, PVs are both important and difficult to learn, which makes it essential to be included in the curriculum. However, it is impossible to teach all of them as there are 12,508 PV lemmas in the BNC alone. Therefore, to create a pedagogical list of PVs will pave the way for learners to improve English proficiency. According to Liu (2012) [11], the 150 most frequently-used PVs compiled in his studies cover 62.95% of the 512,305 PVs in total, which is obviously an efficient pedagogical list of PVs.

Manuscript received May 9, 2019; revised July 25, 2019.

Wenting Yang is with Trinity College Dublin , Ireland (e-mail: weyang@tcd.ie).

doi: 10.18178/ijlll.2019.5.3.218

III. Difficulties and Benefits of Using this Type of Corpus

The TC is a kind of textbook corpora, allowing English learners to search example sentences of particular phrases or words shown in the corpora. Nevertheless, creating these types of corpora is rather time-consuming. Take TC as an example, three steps are involved when creating the corpus – transcription, proofreading and part-of speech tagging. Firstly, no editable electronic textbooks could be found online due to China's Intellectual Property Law, so the author has to transcribe all the passages manually, and then spend a large amount of time on proofreading. Secondly, although SketchEngine automatically annotated passages, there are also minor mistakes, so the vertical file has to be downloaded from the SketchEngine for proofreading part-of-speech tags. The above-mentioned are not tricky problems but rather time-consuming, so efficient technical solutions are required to save corpus researchers the troubles of repetitious manual work. In contemporary society, OCR can help with the transcription process, but errors occur from time to time and manual work is still essential.

Additionally, according to Yoon & Hirvela (2004) [14], other difficulties of using corpus in routine teaching include (1) some learners may encounter difficulties in acquiring the skills needed to experience textbook corpora, and not all English learners are able to gain access to the technology necessary for using the corpus; (2) some concordance programs are particularly sophisticated and produce languages that are difficult to interpret, so some English learners may feel confused in the face of the complex-looking linguistic input. Even though students can be trained to use the corpus, Cobb (1997) [15] claimed that "the amount of time necessary for students to become accustomed to the new technology could well be spent on more conventional and time-tested practices".

Nevertheless, corpus-based language learning also demonstrates many benefits. Thurston & Candlin (1998) [16] claimed, "participants reacted positively toward corpus-based vocabulary teaching, though some negative reactions were observed." Moreover, textbook corpora provide researchers a quantitative approach to compare "school English" and "real word English" and to bridge the gap between the two, which is also the aim of this study.

IV. Information of the TC and the 150 PVS

The corpus is based on the English textbooks published by People's Education Press (PEP). PEP publishes 11 textbooks in total for high school ESL students of different levels. While learning books 1-5 is compulsory, there are no requirements for teachers to include books 6-11 in their teaching routine. Therefore, teachers in the school usually

skip books 6-11 and only focus on books 1-5. In other words, books 1-5 are highly likely to be the only source for high school students to acquire vocabularies. Therefore, to ensure that the result of the study can truly reflect English-teaching materials used at school, the author decides to include only books 1-5 in the corpus. The corpus consists of 21,645 words and 4104 unrepeated words plus punctuations. Although the number of words are consistent with the academic requirements set by the Ministry of Education, such a number of words are unlikely to prepare students to be a proficient English user, because if 98% is an ideal coverage for a non-native speaker to totally understand a material, then an 8000-9000 word-family vocabulary is required for handling the written texts and 6000-7000 families for the spoken text (Gardner & Davies, 2007) [13].

The list of the 150 most frequently-used PVs referred in the study is created by Liu (2012) [17] who aims at comparing the findings of the two previous studies of English PVs by Biber *et al*. (1998) [18] and Gardner & Davies (2007) [13], and creating a new list of PVs based on their studies. The reason for using Liu's (2012) [17] list of PVs as reference is that his list is more comprehensive and concise – he identified 48 additional most-frequently-used PVs and combined two pairs of synonyms (*look around* and *look round*; *turn around* and *turn round*) which were reported as individual PVs in Gardner & Davies' (2007) [13] list of PVs.

V. Method

Firstly, I transcribed all the reading passages in PEP English textbooks 1-5 to Microsoft Word to make it recognizable by the SketchEngine. Then the SketchEngine will automatically annotate those passages.

Secondly, Liu's (2012) [17] list of 150 PVs will be queried in the TC and their frequencies will be tabulated in the excel spreadsheet for comparison. TC is a small corpus which only includes 3490 lexical items in total with only 4.2% of lexicons having a frequency number higher than 20, so it is better to search for lexical items directly and then filter them by tags, rather than query "[lemma]+[tags]". It is because the tags in the TC are generated by SketchEngine automatically, and there may be wrong taggings. Therefore, compare with the time-consuming manual review of annotations, it is more efficient to directly search for the 150 PVs in the small-sized TC.

Thirdly, the frequency of each PV will be converted into a universal unit that allows comparison among corpora of different sizes. One of the most common norming methods used by the researcher to examine word frequencies in corpora of different sizes is "*A number of tokens per number of words*" method. In the study, the number of tokens per million words (PMWs) method will be applied to guarantee the consistency with Liu's (2012) [17] statistics.

Fourthly, after all the data have been tabulated and the raw frequency data have been converted into PMWs, PVs will be analysed in TC, BNC and COCA from four perspectives –frequency, rank order, distribution patterns and coverage percentages.

Rank order differences, frequency differences and coverage differences can be measured by subtraction. For example, as shown in the tables in the appendix, "the rank difference number can be interpreted to mean either that the frequency of *pick up* in COCA is one rank higher (i.e., +1) than its frequency in BNC, or its frequency in BNC is one rank lower (-1) than its frequency in COCA" (Liu, 2011). Moreover, to make the analysis of data simpler, no +/- symbol will be used in the analysis procedures. The same

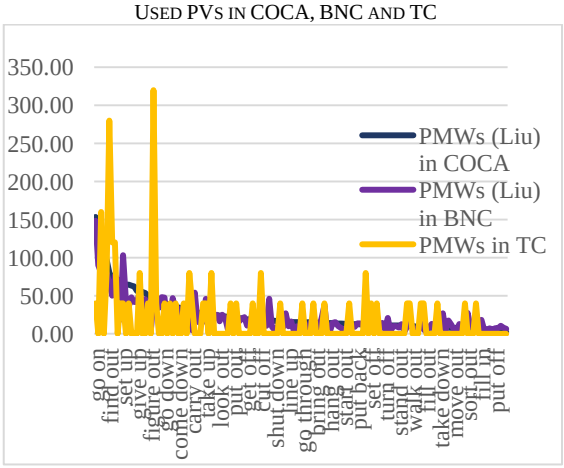
criteria can be applied on the interpreting of differences of frequency and coverage percentage of PVs among different corpora. Nevertheless, though subtraction allows for the calculation of frequency differences and coverage percentage differences, it is not enough for the author to compare the frequency distribution among the three corpora. Therefore, a line graph is essential to make the difference visible.

VI. Findings and Discussion

Only 41 PVs (27.3%) on Liu's (2012) [17] list are found in the TC, but the PVs in TC accountted for 0.3% of the total number of words, which is rather close to those in COCA (0.4%) and BNC (0.3%). However, the frequency distribution pattern of PVs in TC appears to be significantly different from those in BNC and COCA.

Line graphs can be directly applied to make the frequency distribution patterns of PVs visible. As shown in Table I, COCA and BNC have rather similar frequency distributions, and the largest frequency difference value between the two corpora is only 51.12. On the other hand, the 150 PVs in TC do not show any linear correlation with the PV frequency distribution in either COCA or BNC. Moreover, 15 PVs between TC and COCA, and 14 PVs between TC and BNC produce frequency difference values higher than 51.12. Additionally, mean is also an indicator of frequency difference. While the average frequency difference value between TC and COCA, and between TC and BNC are 25.26 and 23.53 respectively, the value between COCA and BNC is 8.52, which altogether show that the frequency differences between TC and the two large corpora are significant.

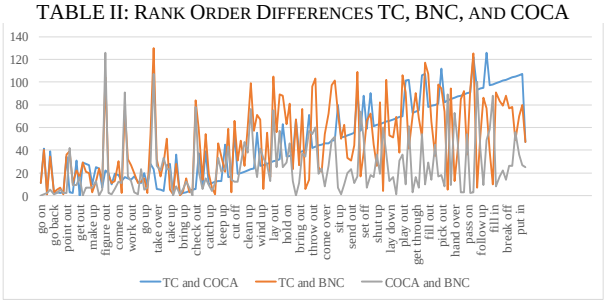
TABLE I: Frequency Distribution of the 150 Most Frequently-Used PVs in COCA, BNC and TC



Additionally, Liu (2012) [17] used two-way chi-square test to determine whether the relative frequency of the PVs was statistically equal in BNC and COCA. The formula is "the total observed frequencies of the 150 PVs measured against the total number of words of their respective corpora minus the total number of tokens of the 150 PVs". Liu (2012) [17] claimed that the difference in corpus size was controlled in this way. Nevertheless, due to the large difference in the size between TC and COCA, and between TC and BNC, Liu's method cannot be used in this study.

Rank order difference is also an important indicator for corpora differences. As shown in Table II, a smaller rank order difference is found between BNC and COCA, with the mean at 25.06, which is almost half the mean of rank differences between TC and BNC (47.2), and between TC and COCA (47.02). When looking at individual data, the PVs' frequency rank orders are relatively identical in BNC and COCA. For example, as is shown in the appendix, 6 PVs (4%) share the same rank orders, and 7 PVs (4.7%) only show a single digit difference (e.g., *go back* ranks 5th in COCA and 4th in BNC, a rank difference of 1). Furthermore, 8 out of 10 PVs in the COCA are also shown in the BNC. Nevertheless, when looking at the two corpora as a whole, 98 (65.3%) PVs record a rank order difference larger than 10. Furthermore, when looking at rank order differences between TC and COCA, and between TC and BNC, the differences are obviously more significant. Firstly, 149 PVs (99.3%) in TC do not share the same rank order with the PVs in either BNC or COCA. Additionally, 128 (85.3%) PVs in BNC and 127 (84.7%) in COCA record a number of rank difference larger than 10 when comparing to the rank order of PVs in the TC.

TABLE II: Rank Order Differences TC, BNC, and COCA



VII. Relevant Case Studies

In the section, the findings of two influential research focusing on identifying frequently-used PVs would be introduced.

A. Liu's (2011) Frequency List

Liu (2011) [11] studied the lists of PVs created by Biber *et al*. (1999) [12] and Gardner and Davies (2007) [13], and discovered that the two lists highly overlap with each other, with only 4 PVs in the list created by Biber *et al*. (1999) [12] are not in Gardner and Davies' (2007) [13] list of 100 PVs. As a result, the list of the most frequently-used PVs was expanded to include 104 PVs.

Additionally, Liu (2011) [11] also searched for other commonly-used PVs in BNC and COCA by using four recent comprehensive PV dictionaries as a searching guide.

Liu (2011) [11] searched for 8847 PVs in the corpus, but only 152 PVs can make the list. Nevertheless, those PVs account for 62.95% of the total PV occurrences (512,304) in the BNC though only covers 1.2% of the total PV lemmas in the corpus.

B. Gardner & Davies' (2007) Frequency List

Gardner & Davies (2007) [13] used BNC to query every example of the [V+AVP] combination, and all inflection forms of the same verb were counted together. By doing so,

the researchers found that,

(1) the top 20 lexical verbs found in the PV constructions accounted for 53.7% of all the PVs in the BNC;

(2) these 20 lexical verbs, combined with only 8 particles, making up 50.4% of PVs in the BNC;

(3) 25 PVs account for nearly one-third of all PV occurrence in the corpus and the list of 100 PVs make up 51.4% (Garnier & Schmitt, 2015) [6].

VIII. Conclusion

The purpose of the study is to explore the quality of the PEP English textbooks by measuring if the books can provide students with necessary language materials – e.g., PVs. However, it seems that the PVs chosen in the TC are largely based on intuition instead of scientific quantitative research. It is because the PVs in the TC differed from those in BNC and COCA in the four perspectives measured – frequency, rank order, distribution patterns and coverage percentages. In addition, the TC only covers 27.3% of Liu's (2011) [11] list of PVs, which indicates that the PEP textbooks do not provide enough language sources to prepare students to be proficient English users.

Appendix

	In COCA		In BNC		In TC	
PVs	PMWs (Liu)	Rank Order (Liu)	PMWs (Liu)	Rank Order (Liu)	PMWs	Raw Frequency	Rank Order
go on	153.48	1	148.33	1	39.98	1	12
pick up	115.40	2	89.95	3	0.00	0	42
come back	109.44	3	79.91	5	159.93	4	4
come up	101.48	4	54.97	9	0.00	0	43
go back	97.31	5	80.27	4	79.96	2	7
find out	80.43	6	65.88	8	279.88	7	3
come out	72.51	7	49.99	12	119.95	3	5
go out	70.77	8	76.52	6	119.95	3	5
point out	69.71	9	69.51	7	0.00	0	43
grow up	69.56	10	18.44	52	39.98	1	13
set up	65.11	11	103.12	2	39.98	1	13
turn out	64.58	12	42.64	21	0.00	0	43
get out	64.43	13	35.28	29	39.98	1	13
come in	63.36	14	47.91	14	0.00	0	43
take on	62.17	15	41.79	22	0.00	0	43
give up	56.11	16	41.66	23	0.00	0	43
make up	55.80	17	54.43	10	79.96	2	7
end up	54.80	18	33.62	31	0.00	0	43
get back	53.56	19	45.31	19	0.00	0	43
look up	50.24	20	38.53	25	39.98	1	13
figure out	48.17	21	2.73	147	0.00	0	43
sit down	47.43	22	44.57	20	319.86	8	2
get up	47.41	23	39.18	24	39.98	1	13
take out	44.32	24	34.10	30	0.00	0	43
come on	43.22	25	48.07	13	0.00	0	43
go down	39.62	26	47.59	15	39.98	1	13
show up	39.57	27	7.64	118	0.00	0	43
take off	36.58	28	21.52	45	39.98	1	13
work out	36.47	29	46.81	16	0.00	0	43
stand up	36.46	30	30.43	33	39.98	1	13
come down	34.58	31	32.90	32	0.00	0	43
go ahead	33.80	32	17.47	55	0.00	0	43
go up	33.23	33	36.61	28	39.98	1	13
look back	29.97	34	22.40	41	0.00	0	43
wake up	29.54	35	16.07	61	79.96	2	7
carry out	28.86	36	4.60	143	39.98	1	13
take over	28.24	37	53.95	11	0.00	0	43
hold up	28.16	38	16.16	60	0.00	0	43
pull out	27.42	39	13.99	72	0.00	0	43
turn a/round	27.37	40	15.62	63	39.98	1	13
take up	27.19	41	45.86	18	39.98	1	13
look down	24.96	42	22.11	42	0.00	0	43
put up	24.49	43	28.22	35	79.96	2	7
bring back	24.34	44	21.90	43	0.00	0	43
bring up	24.31	45	24.95	39	0.00	0	43
look out	23.97	46	16.33	58	0.00	0	43
bring in	23.23	47	24.93	40	0.00	0	43
open up	22.74	48	20.43	48	0.00	0	43
check out	22.35	49	5.73	127	0.00	0	43
move on	21.18	50	14.12	71	39.98	1	13
put out	21.07	51	16.52	57	0.00	0	43
look a/round	20.75	52	14.67	67	39.98	1	13
catch up	20.39	53	16.05	62	0.00	0	43
go in	20.37	54	19.65	50	0.00	0	43
break down	19.15	55	21.89	44	0.00	0	43
get off	18.85	56	10.81	89	0.00	0	43
keep up	18.85	56	13.38	77	0.00	0	43
put down	18.75	58	28.60	34	39.98	1	13
reach out	18.74	59	9.45	102	0.00	0	43
go off	18.62	60	20.94	46	0.00	0	43
cut off	18.01	61	13.74	73	79.96	2	7
turn back	17.91	62	13.67	74	0.00	0	43
pull up	17.81	63	10.53	91	0.00	0	43
set out	17.67	64	46.11	17	0.00	0	43
clean up	17.58	65	9.26	103	0.00	0	43
shut down	16.92	66	4.69	142	0.00	0	43
turn over	16.50	67	9.70	100	0.00	0	43
slow down	16.29	68	11.91	84	39.98	1	13
wind up	16.02	69	8.26	110	0.00	0	43
turn up	15.62	70	26.97	37	0.00	0	43
line up	15.51	71	9.96	98	0.00	0	43
take back	15.47	72	16.20	59	0.00	0	43
lay out	15.27	73	2.64	148	0.00	0	43
go over	15.25	74	9.86	99	0.00	0	43
hang up	15.23	75	5.40	132	0.00	0	43
go through	15.22	76	9.66	101	39.98	1	13
hold on	15.19	77	9.04	106	0.00	0	43
pay off	15.09	78	6.18	124	0.00	0	43
hold out	15.06	79	15.00	66	0.00	0	43
break up	14.91	80	12.80	80	39.98	1	13
bring out	14.53	81	14.18	70	0.00	0	43
pull back	14.47	82	7.50	119	0.00	0	43
hang on	14.34	83	20.11	49	0.00	0	43
build up	14.21	84	37.34	27	39.98	1	13
throw out	14.13	85	4.91	139	0.00	0	43
hang out	14.10	86	2.74	146	0.00	0	43
put on	14.08	87	14.21	68	0.00	0	43
get down	13.53	88	15.31	64	0.00	0	43
come over	13.43	89	9.99	97	0.00	0	43
move in	13.43	89	7.86	115	0.00	0	43
start out	13.14	91	4.88	140	0.00	0	43
call out	13.03	92	3.79	144	0.00	0	43
sit up	12.83	93	11.53	86	39.98	1	13
turn down	12.71	94	10.46	93	0.00	0	43
back up	12.58	95	9.08	105	0.00	0	43
put back	12.53	96	13.63	76	0.00	0	43
send out	12.40	97	13.67	74	0.00	0	43
get in	12.36	98	11.22	87	0.00	0	43
blow up	12.11	99	7.79	116	79.96	2	7
carry on	12.04	100	38.51	26	0.00	0	43
set off	11.79	101	18.60	51	39.98	1	13
keep on	11.71	102	8.28	109	0.00	0	43
run out	11.57	103	11.89	85	39.98	1	13
make out	11.35	104	11.00	88	0.00	0	43
shut up	11.27	105	14.19	69	0.00	0	43
turn off	11.25	106	5.91	125	0.00	0	43
bring about	11.22	107	20.73	47	0.00	0	43
step back	10.92	108	3.31	145	0.00	0	43
lay down	10.89	109	10.08	96	0.00	0	43
bring down	10.86	110	10.17	94	0.00	0	43
stand out	10.85	111	8.14	112	0.00	0	43
come along	10.68	112	12.64	81	0.00	0	43
play out	10.32	113	2.62	149	0.00	0	43
break out	10.26	114	9.11	104	39.98	1	13
go a/round	10.07	115	17.96	54	39.98	1	13
walk out	10.01	116	8.06	113	0.00	0	43
get through	9.70	117	5.31	133	0.00	0	43
hold back	9.16	118	8.19	111	0.00	0	43
write down	9.04	119	15.05	65	39.98	1	13
move back	8.63	120	5.63	130	39.98	1	13
fill out	8.46	121	2.55	150	0.00	0	43
sit back	8.43	122	8.30	108	0.00	0	43
rule out	8.25	123	13.05	78	0.00	0	43
move up	8.21	124	4.75	141	0.00	0	43
pick out	8.19	125	8.52	107	39.98	1	13
take down	8.19	125	7.71	117	0.00	0	43
get on	8.17	127	26.83	38	0.00	0	43
give back	7.97	128	5.05	137	0.00	0	43
hand over	7.96	129	17.35	56	0.00	0	43
sum up	7.77	130	12.28	83	0.00	0	43
move out	7.76	131	5.70	128	0.00	0	43
come off	7.67	132	5.13	135	0.00	0	43
pass on	7.42	133	12.81	79	0.00	0	43
take in	7.07	134	5.07	136	0.00	0	43
set down	6.95	135	5.02	138	39.98	1	13
sort out	6.82	136	27.36	36	0.00	0	43
follow up	6.73	137	10.12	95	0.00	0	43
come through	6.66	138	5.64	129	0.00	0	43
settle down	6.53	139	10.76	90	39.98	1	13
come a/round	6.50	140	12.42	82	0.00	0	43
fill in	5.99	141	18.18	53	0.00	0	43
give out	5.62	142	5.30	134	0.00	0	43
give in	5.58	143	5.76	126	0.00	0	43
go along	5.28	144	7.14	122	0.00	0	43
break off	4.77	145	5.46	131	0.00	0	43
put off	4.67	146	7.39	120	0.00	0	43
come about	4.63	147	7.38	121	0.00	0	43
close down	4.13	148	10.48	92	0.00	0	43
put in	4.00	149	8.06	113	0.00	0	43
set about	2.32	150	6.42	123	0.00	0	43

References

[1] A. Cornell, "Realistic goals in teaching and learning phrasal verbs," *IRAL-International Review of Applied Linguistics in Language Teaching*, vol. 23, no. 1-4, pp. 269-280, 1985.

[2] A. Kurtyka, "Teaching English phrasal verbs: A cognitive approach," *Applied Cognitive Linguistics II: Language Pedagogy*, pp. 29-54, 2001.

[3] G. Kartal, "Phrasal verbs in ELT coursebooks used in Turkey: A corpus-based analysis," *Cumhuriyet International Journal of Education*, vol. 7, no. 4, pp. 534-550, 2018.

[4] J. Chen, "On how to solve the problem of the avoidance of phrasal verbs in the Chinese context," *International Education Journal*, vol. 8, no. 2, pp. 348-353, 2007.

[5] N. Condon, "How cognitive linguistic motivations influence the learning of phrasal verbs," *Applications of Cognitive Linguistics*, vol. 6, p. 133, 2008.

[6] M. Garnier and N. J. Schmitt, "The PHaVE list: A pedagogical list of phrasal verbs and their most frequent meaning senses," *Language Teaching Research*, vol. 19, no. 6, pp. 645-666, 2015.

[7] Y. Liao and Y. J. Fukuya, "Avoidance of phrasal verbs: The case of Chinese learners of English," *Language Learning*, vol. 54, no. 2, pp. 193-226, 2004.

[8] J. H. Hulstijn and E. Marchena, "Avoidance: Grammatical or semantic causes?" *Studies in Second Language Acquisition*, vol. 11, no. 3, pp. 241-255, 1989.

[9] C. M. Darwin and L. S. Gray, "Going after the phrasal verb: An alternative approach to classification," *TESOL Quarterly*, vol. 33, no. 1, pp. 65-83, 1999.

[10] M. McCarthy, *Touchstone from Corpus to Course Book*, pp. 1-20, 2004.

[11] D. Liu, "The most frequently used English phrasal verbs in American and British English: A multicorpus examination," *TESOL Quarterly*, vol. 45, no. 4, pp. 661-688, 2011.

[12] D. Biber, S. Johansson, G. Leech, S. Conrad, and E. Finegan, *Longman Grammar of Spoken and Written English*, Pearson Education Ltd., pp. 1999-1204, 1999.

[13] D. Gardner and M. Davies, "Pointing out frequent phrasal verbs: A corpus-based analysis," *TESOL Quarterly*, vol. 41, no. 2, pp. 339-359, 2007.

[14] H. Yoon and A. Hirvela, "ESL student attitudes toward corpus use in L2 writing," *Journal of Second Language Writing*, vol. 13, no. 4, pp. 257-283, 2004.

[15] T. Cobb, "Is there any measurable learning from hands-on concordancing?" *System*, vol. 25, no. 3, pp. 301-315, 1997.

[16] J. Thurston and C. N. Candlin, "Concordancing and the teaching of the vocabulary of academic English," *English for Specific Purposes*, vol. 17, no. 3, pp. 267-280, 1998.
[17] D. Liu, "The most frequently-used multi-word constructions in academic written English: A multi-corpus study," *English for Specific Purposes*, vol. 31, no. 1, pp. 25-35, 2012.
[18] D. Biber, B. Douglas, S. Conrad, and R. Reppen, *Corpus linguistics: Investigating Language Structure and Use*, Cambridge University Press, 1998.

Wenting Yang is from China, she is currently a M.Phil student majoring in linguistics in Trinity College Dublin.

She has 5 years of English teaching experience and has published one essay titled *Challenges for Traditional Teachers in MOOCs Era.* Her research interests include second LANGUAGE acquisition and teaching material development.