The Most Frequently Used English Phrasal Verbs in American and British English: A Multicorpus Examination

DILIN LIU
University of Alabama
Tuscaloosa, Alabama, United States

This study uses the Corpus of Contemporary American English and the British National Corpus as data and Biber, Johansson, Leech, Conrad, and Finegan's (1999) and Gardner and Davies' (2007) informative studies as a starting point and reference. The study offers a cross-English variety and cross-register examination of the use of English phrasal verbs (PVs), one of the most difficult aspects of English for learners of English as a foreign language or English as a second language. The study first identified the frequency and usage patterns of the most common PVs in the two corpora and then analyzed the results using statistical procedures, the chi-square and dispersion tests, to determine any significant cross-variety or -register differences. Besides validating many of the findings of the two previous studies (although neither was a cross-English variety examination), the results of this study provide new, useful information about the use of PVs. In addition, the study resulted in a comprehensive list of the most common PVs in American and British English, one that complements those offered by the two previous studies with more necessary items and more detailed usage information. The study also presents a cross-register list of the most frequent PVs, showing in which register(s) each of the PVs is primarily used. Finally, pedagogical and research implications are discussed.
doi: 10.5054/tq.2011.247707

Because of their extremely high frequency in the English language and the great difficulty they present to language learners, phrasal verbs (PVs) have long been a subject of interest and importance in English as a foreign language (EFL) or English as a second language (ESL) teaching and research, as evidenced by the many publications on the topic (Bolinger, 1971; Cordon & Kelly, 2002; Darwin & Gray, 1999; Gardner & Davies, 2007; Liao & Fukuya, 2004; McCarthy & O'Dell, 2004; Side, 1990; Wyss, 2003). The unique challenge for teaching PVs is that, although PVs are ubiquitous in the English language, EFL or ESL

speakers, especially those with a lower and intermediate level of proficiency, consistently avoid using them (Dagut & Laufer, 1985; Hulstijn & Marchena, 1989; Laufer & Eliasson, 1993; Liao & Fukuya, 2004). The reasons for this avoidance are many, including cross-linguistic differences and the complexity of syntactic and semantic structures of PVs (Dagut & Laufer, 1985; Hulstijn & Marchena, 1989; Laufer & Eliasson, 1993). The enormous number of PVs in English also contributes to the problem, because it makes learners feel overwhelmed, not knowing which ones to learn. Thus identifying the most useful PVs is paramount for language learning purposes. Although the answer to the question of which PVs are useful may vary depending on learners' objectives and learning contexts, frequency is usually a good criterion for determining usefulness. This is because, in general, highly frequent PVs are more useful than those with very low frequency. There have been two corpus-based frequency studies of English PVs (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Gardner & Davies, 2007), and both have provided valuable information about PVs and their distribution patterns. Yet, there are important limitations in each of the two studies. It is important, however, to point out that the limitations are not due to any oversight on the part of the scholars who did the studies but simply the result of their specific foci and space constraints.

Being a small section of a comprehensive book on English grammar, Biber et al.'s (1999) treatment of PVs is limited largely to a small set of PVs (31 in total). Gardner and Davies' (2007) work, though covering many more PVs than Biber et al.'s work, has three limitations of its own. First, their list of the most frequent PVs (a total of 100 items) contains only PVs made up of the top 20 PV-producing lexical verbs (e.g., *come, go, get,* and *take*). In other words, the list does not include highly frequent PVs formed by verbs outside the top 20 PV-producing ones (e.g., *keep up* is not on the list because *keep* is not one of the top 20 PV-producing verbs). As a result, their study, although offering new insights about PVs (e.g., a very small group of lexical verbs make up a majority of PVs), does not provide a thorough account of the most frequent PVs. Second, with the British National Corpus (BNC) as the data source, their study deals exclusively with British English. It remains an interesting question whether their findings are also true of any other major varieties of English. In fact, in their conclusion, Gardner and Davies themselves explicitly called for the need to test the validity of their list "against other megacorpora" (p. 354). Third, limited by space, their study did not render a cross-register examination of the frequently used PVs. Such cross-register information is, however, very important for language learning purposes, because it indicates the contexts where specific PVs are and are not typical. Gardner and Davies also explicitly recommended "a reanalysis of the [PV] lists across major registers (e.g., spoken versus

written English)" (p. 354). In order to help fill in the aforementioned information gaps about PVs, the present study aims to offer a comparative investigation of the most frequently used PVs between American and British English and an examination of the usage information of these frequently used PVs across registers in American English.

DEFINITION OF PHRASAL VERB

For any study of PVs, the definition of PV is often the first order of business. Yet, what constitutes a PV and how to classify PVs have long been topics of debate. Many different theories have been proposed, and they differ largely over what syntactic and semantic features define a PV and how such features should be used to classify PVs (Biber et al., 1999; Celce-Murcia & Larsen-Freeman, 1999; Darwin & Gray, 1999; Gardner & Davies, 2007; Quirk, Greenbaum, Leech, & Svartvik, 1985). However, many of the differences among the theories are quite minuscule, especially from a language learner's perspective. As Gardner and Davies (2007, p. 341) correctly note, "if even the linguists and grammarians struggle with nuances of PV definitions, of what instructional value could such distinctions be for the average second language learner?" Furthermore, because of the purposes of the present study, there is little need and room for a lengthy review of the various definitions that have been proposed so far. This study had two main purposes: (1) to examine in the Corpus of Contemporary American English (COCA) the frequencies of the most common PVs and to compare the results with those reported in Biber et al. (1999) and Gardner and Davies (2007); and (2) to conduct a cross-register distribution analysis of the PVs in COCA and to compare the results with those of the study by Biber et al.

In order to ensure a meaningful comparison between the findings of this study and those of the other two, this study uses Gardner and Davies' (2007) definition of VP: any two-part verb "consisting of a lexical verb (LV) proper . . . followed by an adverbial particle (tagged as AVP) that is either contiguous (adjacent) to that verb or noncontiguous (i.e., separated by one or more intervening words)" (p. 341). The reason for using Gardner and Davies' definition rather than Biber et al.'s is twofold. First, it is simpler, because it involves only one syntactic criterion: "a verb plus an AVP." In contrast, Biber et al.'s definition includes an additional semantic component: PVs must "have meanings beyond the separate meanings of the two parts [i.e., the verb and the AVP]" as in the case of "*come on, shut up* . . ." whereas verb + AVP combinations in which "the verb and the adverb have their own meanings" are "free combinations like *come back, come down* . . ." (Biber

et al., 1999, p. 404). The application of this semantic criterion is not always straightforward and often involves some subjective judgments. Of course, Gardner and Davies' syntactic criterion is not always simple either, because whether a verb particle should be classified as an AVP, regular adverb, or preposition is sometimes open to debate, an issue I address later. The second reason for using Gardner and Davies' definition is that, as is shown next, a majority of the most frequent PVs examined in this study came from Gardner and Davies' study.

METHOD

Corpora Used

As mentioned earlier, the main corpus used for this study was COCA, a large free online corpus developed by Professor Mark Davies of Brigham Young University. When this study was conducted, COCA consisted of 386.89 million words via data gathered from 1990 to 2008, that is, an average of approximately 20 million words from each of the 19 years. The corpus contains five subcorpora: spoken, fiction, magazine, newspaper, and academic writing, with each subcorpus contributing an equal amount of data (4 million words per subcorpus per year). The corpus is also user friendly. Its search engine allows the user to perform, among other things, the search and comparison of "the frequency of words, phrases and grammatical constructions" (Davies, 2008). Besides COCA, the 100.47-million-word BNC was also used both indirectly and directly: The frequency results of the 100 most common PVs in the BNC reported in Gardner and Davies' study were compared with the PVs' frequencies in COCA, and I queried the BNC directly through Davies' (2005) BYU interface for the frequency information of the other PVs that are not on Gardener and Davies' list of the 100 most frequent PVs. Furthermore, because the results of Biber et al.'s study were also used for comparison in this study, the corpus they used, the 40-million-word Longman Spoken and Written English (LSWE) corpus, was also indirectly used in this study.

To help the reader better understand the cross-corpora comparisons to be rendered in the Findings and Discussion section, some relevant background information about the LSWE and the BNC is given here. The spoken part of the LSWE consists primarily of face-to-face conversation (see Biber et al., 1999, p. 29–30). Similarly, a very large portion of the spoken subcorpus of the BNC is composed of such conversations. In fact, the British English portion of the LSWE is included in the BNC. In contrast, the spoken part of COCA consists mostly of TV or radio broadcasting speech.

Data Gathering and Data Reporting or Analysis Methods

Querying for the frequency of a PV is a challenging task. One cannot accomplish the search by simply entering the lexical verb lemma of a PV in the form of [verb] plus its particle (e.g., "[go] on"), because not every one of the tokens generated by such a search is a phrasal verb. For example, the "[go] on" entry may yield non-PV tokens such as "We typically go on Mondays" where "on" is a preposition in the time adverbial phrase "on Mondays," not an adverbial particle (AVP) of *go*. (The lemma search function helps generate the tokens of the various forms of the verb, e.g., *go/goes/going/went/gone* for the lemma *go*.) Thus, to ensure an accurate count of all the tokens of a PV, sophisticated query methods are called for. One such method is found in Gardner and Davies' (2007) study. They imported the entire tagged BNC data set into the Microsoft SQL server, a relational data program that can help identify all the instances of PVs. This method was not used in this study, however, because COCA does not make its entire tagged data set accessible to the public. Instead, this study employed basically a four-step procedure using the existing search functions in COCA's interface. This procedure, though more labor intensive, proved to be functional and fundamentally accurate.

The first step was the search for all the PV tokens of a lexical lemma. This was done by entering the verb lemma in the form of [verb] plus [RP*] (RP is the search code for AVPs in COCA and the wildcard * stands for any AVPs). For example, for all the PV tokens of the lexical verb lemma [go], "[go] [RP*]" is entered. The query will generate all the "*go* plus AVP" PV tokens, including *go on, go off,* and so forth. The second step was a search of the tokens of transitive PVs used with their AVPs separated by one intervening word. This was carried out by entering for search "[verb] * [RP*], with the wildcard * between the verb and the AVP standing for any intervening word. The third step was the search of the tokens of separable PVs with two intervening words (e.g., *look the word up*). This task was performed by entering "[verb] * * [RP*]". No search was done, however, for instances of PVs with their AVPs separated by three or more intervening words. This is because PVs so used are rare, and a search for them will yield "many false PVs" (Gardner & Davies, 2007, pp. 344–345). Furthermore, Gardner and Davies did not include such tokens, making it necessary to exclude them in this study to ensure a meaningful comparison. In steps 2 and 3, I read through the result lines to exclude any false tokens. All the aforementioned searches were performed with the cross-section comparison search function in COCA activated so that the search results included the PVs' frequency distribution in each of the five registers. The last step was the recording and tabulation of the query results, using Excel spreadsheets. For each PV, the frequency results of its various forms in the five registers were entered, and the subtotal and total frequencies were computed.

As far as the frequency counting or reporting method is concerned, raw frequency numbers cannot be used for comparison purposes, because of the large differences in size among the corpora used in the study. Instead, a *number of tokens per number of words* norming method must be employed. For examining data in large corpora, researchers typically use the number of tokens per million words (PMWs) method (cf. Biber, Conrad, & Reppen, 1998; Biber et al., 1999; Liu, 2003, 2008; Moon, 1998). Furthermore, given that this method was already used in Biber et al. (1999), it was adopted for this study for the reporting of most of the data. However, in the statistical analysis (i.e., the chi-square and the dispersion tests) of the results to determine whether there were significant differences among the PVs' distributions, I used only raw observed frequencies, because normalized data are inappropriate for such statistical tests.

PVs Examined

In order to render a comparison of the results of this study with those of Biber et al. (1999) and Gardner and Davies (2007), I queried COCA for the frequency of all the PVs in their lists. There were a total of 31 PVs in Biber et al. Each had at least 40 tokens PMWs in at least one register of the LSWE. Gardner and Davies' list consists of 100 items made up of the top 20 PV-producing verbs. Twenty seven of the 31 in Biber et al.'s list overlap, however, with those in Gardner and Davies' list. In other words, only 4 of Biber et al's 31 PVs are not in Gardner and Davies' list. Of these four, one is *go ahead*. It is not in Gardner and Davies' list because *ahead* is not tagged as an AVP in the BNC (or in COCA), but rather it is tagged as a regular adverb. The other three PVs not on Gardner and Davies' list are *shut up, stand up,* and *run out* because *run, shut,* and *stand* are not among the top 20 PV-producing lexical verbs that Gardner and Davies identified. Because of the overlapping of 27 items, the total number of PVs from Biber et al.'s and Gardner and Davies' studies was 104, not 131.

Besides searching these 104 PVs in COCA, I also queried the COCA and the BNC for the other most common PVs. To do so, I used the four most recent comprehensive PV dictionaries as a search list guide: *Cambridge International Dictionary of Phrasal Verbs* (1997), with over 4,500 entries; *Longman Phrasal Verbs Dictionary* (2000), with over 5,000 PVs; *NTC's Dictionary of Phrasal Verbs and Other Idiomatic Verbal Phrases* compiled by Spears (1993), with 7,634 entries; and *Oxford Phrasal Verbs Dictionary for Learners of English* (2001), with over 6,000 entries. I searched a total of 8,847 PVs, 5,933 of which were from the dictionaries, whereas 2,914 were not. The latter were not searched intentionally but were the by-product of my query method [verb] [RP*] which would automatically return all the PVs of the verb being queried, including those not in the dictionaries. For

example, my queries [drive] [RP*] returned not only the intended PVs from the dictionaries, for example, *drive away/up/down/off,* but also those not listed in the dictionaries, for example, *drive about/along/by/round.* Considering the large number of PVs listed in each of the four dictionaries, one may wonder why only 5,933 PVs were queried. The reasons were (1) many of the entries in the dictionaries overlap, and (2) the dictionaries include *verb + preposition* structures (e.g., abide *by* and accede *to*) that are not considered PVs relative to the definition used in this study.

According to Gardner and Davies' (2007) search, there are a total of 12,508 PV lemmas in the BNC. This means that my query of 8,847 left 3,661 PVs unsearched. This should not, however, be a concern for the following reasons. First, the purpose of my study was to identify the most frequently used PVs, and the criterion for inclusion in my list was 10 tokens PMWs. As the immediately following discussion shows, only 152 out of the 8,847 made the list. Most PVs simply do not have the required frequency. Second, my search covered all the lexical verb lemmas that had a total of 1,000 tokens in the BNC or 3,869 in COCA, because this was the minimum number that would give the verbs the potential for yielding the required number of PV tokens to make the most common PV list. Finally, because of tagging errors, not all of the 12,508 PV lemmas are PVs.

As already stated, the criterion for a PV to make the most frequently used list in this study was 10 tokens PMWs in either COCA or the BNC. The rationale for using this criterion was threefold. First, 73 (70%) of the 104 PVs on the Biber et al. and Gardner and Davies' combined list each have 10 tokens or more PMWs; only 31 on Gardner and Davies' list each show a frequency fewer than 10 PMWs. Second, in order to be truly meaningful, a list of the most frequently used PVs should not be too long. Third, as Gardner and Davies (2007) reported, the 100 frequently used PVs they identified already "account for more than half (51.4%) of all the PV occurrences in the BNC" (p. 351).[1] Using this ten-token PMWs criterion, my search identified 48 additional most frequently-used PVs. The search results also showed that these 48 PVs and the four from

[1] It is necessary to note that there is an error in the frequency number of a PV in Gardner and Davies' data that has an implication for the total numbers they reported. In their 100 most common PV list, *carry out* is ranked as the 2nd most frequent PV, boasting a frequency of 10,798. This frequency number is unusually high and incorrect, based on my search and consultation with Mark Davies, one of the authors of the Gardner and Davies article. The correct number is 4,180, which means that their reported frequency of this PV is 6,618 tokens over the actual frequency. This should also have resulted in an inflation of the total PV occurrences in the BNC by 6,618. Thus, with the 6,618 removed from both the token numbers of the 100 PVs (266,926 − 6,618) and the total token numbers of all the PV occurrences in the BNC (518,283 − 6,618), the tokens of the 100 PVs (260,168) should account for 50.78%, instead of the 51.7%, of the total PV tokens (512,305) in the BNC. These adjusted correct numbers are used in the discussion in the remainder of the article. Also, in the appendix, the frequency number and order of *carry out* in the BNC list is adjusted accordingly (from 2nd to 24th).

Biber et al. that are not on Gardner and Davies' list together account for another 12.17% of all the PV occurrences in the BNC. This means that the 152 most frequently-used PVs compiled in this study, while comprising only 1.2% of the total 12,508 PV lemmas in the BNC, cover 62.95% of all the total 512,305 PV occurrences. This helps demonstrate the representativeness and hence the usefulness of these most-frequently used PVs. Of course, there are several limitations that should be considered when using this list for learning/teaching purposes, such as the fact that it is a lemmatized list and that many of the PVs have multi-meanings, two very important issues I will address in the next section.

FINDINGS AND DISCUSSION

Most Frequently Used Phrasal Verbs: American English Versus British English

This study has uncovered the frequency information of 152 PVs, including the 100 from the Gardner and Davies list, the four from Biber et al. that are not in Gardner and Davies' list,[2] and the 48 additional most frequent PVs this study has identified. The frequency information is reported in a table format in the appendix, with the PVs listed in order of their frequency in COCA. To allow for an easy comparison of the PVs' frequency in COCA with their frequency in the BNC, their frequency and rank order information in the BNC is also provided (in the second and third columns from the right). It is necessary to note that the total number of PVs in the appendix is 150, not 152, because I combined the PVs in each of the following two related pairs that were reported as individual PVs in Gardner and Davies' study (2007): *look around* and *look round*; *turn around* and *turn round*. Gardner and Davies also have *come round* and *go round* on their list but not *come around* and *go around*, given that the latter forms are the dominant uses in American English, I have included and combined them with the former in this study. The reason for combining the two forms in each pair is that they are synonymous and that they represent mainly a usage variation between American and British English, an issue that is discussed later.

Before proceeding to a detailed comparison of the PVs' frequency and usage patterns in the two corpora, I briefly discuss how some of the results of this study support Biber et al.'s (1999) and Gardner and Davies' (2007) findings about an interesting aspect of PVs: A relatively small number of lexical verbs and AVPs form the majority of the PVs in English. Biber et al.

[2] One of them is *go ahead*. Even though it is not tagged as a PV, as mentioned earlier, I have included it not only because Biber et al. (1999) did but also because I believe *ahead* is actually an AVP for the verb *go*, making the phrase a true PV.

identified eight verbs and six adverbs as the most productive in forming PVs. Gardner and Davies identified the top 20 PV-producing verbs and the four most "prolific" AVPs that help form for more than half (53.7%) of all the PVs in the BNC (2007, p. 347).[3] The same pattern is found in the lexical verbs and the AVPs in the 52 additional most frequent PVs (48 identified in this study and 4 from Biber et al.). For example, *out* and *up* are each the AVPs in 19 of the 52 PVs, that is, they combine for the AVPS of 38 (73.08%) of the 52 PVs. Concerning the verbs in these 52 PVs, it is important to first recall that all of them are outside the top 20 PV-producing lexical verbs. Yet even these less productive verbs show some concentrated use in PVs. One of them (*hang*) appears in three of the 52 PVs, and five (*fill, keep, pull, show, stand*) each appear in two.

To compare the PVs' frequency distribution patterns in the two corpora, it is necessary to note that the data of the two corpora do not come from the same time period. Although the BNC covers the 1980s to 1993, COCA extends from 1990 to the present, that is, COCA starts basically where the BNC ends. This difference in time periods could be responsible for some of the PV usage variations between the two corpora, which is discussed later.[4] To compare the general frequency patterns of the PVs in the two corpora and to determine whether there is any significant difference calls for a chi-square test of the raw observed frequencies. Given the large difference in size between the two corpora, a one-way chi-square test of the observed frequencies of the PVs from the two corpora would not make sense. To account for the effect of the difference in corpus size, I opted for a two-way chi-square test with 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. In this way, the problem of difference in corpus size was controlled, allowing the chi-square test to determine whether the relative frequency of the PVs was statistically equal in both corpora. The results are reported in Table 1 where I also include at the bottom the PVs' frequencies PMWs in the two corpora for easier comparison.

A close look at the test results indicates that, although there is a significant difference between the frequencies of the PVs in the two

[3] Although most of the top PV-producing verbs and AVPs identified by Biber et al. (1999) overlap with Gardner and Davies' (2007), the rank orders of the items between the two lists differ. For example, whereas *take* and *get* are first and second on the Biber et al. list, *go* and *come* are the first two on Gardner and Davies' list (also my COCA list). The difference appears to have resulted from the different definitions of PV used. As mentioned earlier, Biber et al.'s definition involves a semantic criterion, which excludes verb + adverb combinations where verb and AVP hold separate instead of combined meanings. Thus Biber et al. excluded many of the highly frequent PVs formed by *come* and *go* (e.g., *go back* and *come in*) listed in Gardner and Davies (2007).

[4] I owe this idea to an anonymous reviewer, who suggested that the increased use of certain PVs over the past 20 years in COCA may explain their higher frequencies in COCA than in the BNC.

TABLE 1
Comparison of the Most Common PVs' Overall Frequency Patterns in COCA and the BNC

	COCA	BNC	*df*	Chi-square (χ^2)	*P*	Cramer's V
Total observed frequency of the 150 PVs	1,424,836 (+2.7%)*	322,517 (−10.5%)*	1	4,988.65	0.0001	0.0032
Total number of words minus the 150 PVs' total tokens	385,465,164	100,147,483	1	4,988.65	0.0001	0.0032
Frequency PMWs of the 150 PVs	3,682.79	3,210.09				

Note. COCA = Corpus of Contemporary American English; BNC = British National Corpus; PVs = phrasal verbs. *Percentage that the observed frequency deviated from the expected frequency.

corpora, the difference is actually minuscule, as evidenced by the very small *effect size*, a Cramer's V of only 0.0032, and also by the percentages of deviations (PDs) of the observed frequencies from the expected frequencies, with the frequency in COCA being merely 2.7% higher than expected and the frequency in the BNC being only 10.5% lower than expected. The effect size is extremely important for statistical analysis in corpus research, because, as Gries (2010, p. 286) explained, "the large sample sizes that many contemporary corpora provide basically guarantee that even minuscule effects will be highly significant." Thus the significant difference shown by the chi-square test is very likely the result of the large size of the two corpora. Furthermore, a comparison of all the individual PVs' frequency rank order in COCA against their rank order in the BNC (the results reported in the last column of the appendix) indicates that the PVs' frequency rank orders in the two corpora are fairly similar. For example, for each of the following five PVs, its frequency orders in both corpora are identical: *go on* 1st, *come in* 14th, *get back* 19th, *bring back* 44th, and *turn down* 94th. (Incidentally, *go on* is also the most frequent one in Biber et al.'s study.) Eight out of the top 10 PVs in the COCA list also make the top 10 in the BNC list. Forty-six (30.67%) of the 150 PVs show only a single digit difference between their rank orders in the two corpora (e.g., *pick up* ranks 2nd in COCA and 3rd in the BNC, a rank difference of 1).[5] Thirty-seven (24.67%) record a rank order difference between 10 and 19. However, 67 (44.67%) display a rank order difference of 20 or above, an issue I return to later.

[5] This rank difference number can be interpreted to mean, depending on one's perspective, either that the frequency of *pick up* in COCA is one rank higher (i.e., +1) than its frequency in the BNC, or its frequency in the BNC is one rank lower (−1) than its frequency in COCA. To make the reporting of this rank order comparison simpler, no +/− sign is used.

Given the different time periods the BNC and COCA each cover, the absence of a truly large difference in PV use between the two corpora may suggest that PV use has remained fairly stable. This fact may in turn imply that the list of the most frequently used PVs produced in this study may withstand the test of time. In that case, What about the differences in PV uses found between the two corpora, especially the rank disparity of 20 or more found in 67 of the PVs? What might be the cause(s) for the differences? To answer these questions, we should first understand how and to what extent the frequencies and uses of these PVs in the two corpora differ. A close examination reveals that, although the differences of their rank orders between the two corpora offer some interesting information, the difference between a PV's frequencies (numbers of tokens) in the two corpora is a much more informative indicator. For example, the difference between the rank orders of *come up* in the two corpora is only five (4th in COCA and 9th in the BNC), but its frequency difference in the two corpora is 55.45 PMWs. In contrast, *set off* has a rank order difference of 49 but a frequency difference of only 6.81 PMWs. Therefore, I decided to use frequency as the main criterion to examine the individual PVs' distribution differences in the two corpora.

Specifically, I tested for any significant difference between the raw frequencies of those PVs whose frequencies in the two corpora varied by 10 or more PMWs. There were a total of 39 such PVs. Given that the two corpora differ tremendously in size, I conducted a two-way chi-square test employing exactly the same method used for testing the total frequency difference of the 150 PVs in the two corpora reported earlier in Table 1. Because of the large size of the corpora, the chi-square results for the 39 PVs were all significant, but their Cramer's Vs were very small, ranging from 0.0006 to 0.0019. In order to have a shorter and more focused list of PVs which show a truly noticeable difference in their distributions between American and British English, I excluded from the list those PVs with Cramer's Vs lower than 0.001. This resulted in a list of 30 PVs. Twenty are significantly more common in American English: *check out, come out, come up, figure out, get out, go ahead, grow up, hang out, hold up, lay out, pick up, pull out, show up, shut down, take off, end up, turn out, take on, turn a/round,* and *wake up* (cf. the appendix table for their rank orders or frequencies in the two corpora). Ten appear significantly more frequently in British English: *build up, carry on, fill in, get on, set out, set up, sort out, take over, take up,* and *turn up.* Although the reasons for some of the PVs' prominent use in one of the two English varieties are difficult to determine, the causes for some can be attributed to either usage differences between the two varieties of English or the increase of use in American English, for, as mentioned earlier, COCA starts where the BNC ends in terms of the time periods covered.

Regarding usage differences, an examination of some of the tokens of the PVs confirms the following information indicated by some of the PV dictionaries. The significantly larger number of tokens of *fill in* in the BNC appears related to the fact that British English typically uses *fill in* in "*fill in* or *fill something in* a form/document," whereas American English generally uses *fill out* in such cases. The quadrupled use of *check out* in COCA compared to that in the BNC is the result of the multiple functions or meanings of the PV in American English that are not found in British English, such as its meaning "paying for things" at a store and "borrowing items from a library." Furthermore, the far less frequent use of *shut down* in the BNC is mostly due to the fact that, in British English, *shut up* is often used to express the meaning of "closing a business temporarily," a meaning almost always expressed by *shut down* in American English. This fact also helps explain the lower frequency and rank order of *shut up* in COCA.

Another noticeable use difference between American and British English, as mentioned earlier, relates to the use of *around/round* in the PVs such as *come around/round, go around/round, look around/round,* and *turn around/round*. The distribution of *around/round* in these PVs in the two corpora is reported in Table 2. The results demonstrate that, although it is true Americans prefer *around* and British speakers favor *round*, Americans' preference for *around* over *round* is much stronger than the British preference for *round* over *around*. The American use of *around* is more than 90% of the time in each of the four PVs, whereas the British use of *round* is in general much less than 90% of the time.

Concerning frequency differences likely caused by the increased use of certain PVs in American English, a query of COCA indicated that *check out, hang out, show up,* and *come up* each show a noticeable increase in number of tokens from 1990–1994 to 2005–2009. *Check out* increased by 102%, *hang out* by 52%, *show up* by 25%, and *come up* by 23%. Such substantial increases of the PVs in COCA may help explain their higher frequencies in COCA than in the BNC. Yet, because we do not have the British English data after the early 1990s, we cannot be certain whether the same increases would have also occurred in British English. In short, the analysis of the PVs' frequency patterns indicates that, although their general distribution patterns are very similar in both corpora, there are some differences concerning some specific PVs because of (1) usage differences between American and British English and/or (2) increased use in American English. Knowledge of these differences is useful to English language educators when deciding which PVs should be taught and learned in which English variety.

TABLE 2
Distribution of the PV Particles *Around/Round* in COCA and the BNC

PV		COCA		BNC	
Verb	AVP	Tokens PMWs	Percentage	Tokens PMWs	Percentage
come	*around*	6.11	94	1.40	11
	round	0.39	6	11.02	89
go	*around*	9.37	93	4.36	24
	round	0.70	7	13.60	76
look	*around*	20.54	99	7.76	53
	round	0.21	1	6.91	47
turn	*around*	26.82	98	4.21	27
	round	0.55	2	11.41	73

Note. AVP = adverbial particle; PMWs = per million words.

Cross-Register Differences in the Use of PVs

To determine whether there is a significant difference in the overall raw frequency distributions of the 150 PVs among the five registers in COCA, I conducted a one-way chi-square test and a dispersion/adjusted frequency test using Gries' (2008b) Dispersions2 program. This dispersion test yields, in addition to a series of adjusted frequencies, a deviation of proportion (DP) score, which theoretically can range from 0 to 1, but sometimes the number of parts of the corpus and other factors may prevent it from reaching the maximal value of 1. To address this problem, the test also gives a normalized DP score, shown as DP_{norm}, which is able to display the maximal value. The values of DP near 0 suggest that the frequencies of a linguistic item are distributed in proportion to the sizes of the corpus registers or parts, whereas high values, especially those near 1, signify that the frequencies of the linguistic item are distributed very unevenly across the registers. An adjusted frequency is a downwardly adjusted total frequency in proportion to the degree of the unevenness of the distribution of the linguistic item. The results of both the chi-square and the dispersion tests are reported in Table 3. Besides the raw frequencies of the PVs, I have also reported the frequencies PMWs so the results can be compared with those of the Biber et al. (1999) study. The result of the chi-square test is very significant, with $p < 0.0001$, but the deviations of the PVs across the registers are not particularly high according to Gries's DP (0.214; 0.268_{norm}). It is important to note, however, that the specific percentage deviations of the observed frequencies of the PVs from the expected are fairly high: Whereas the observed frequencies in the spoken and fiction registers are, respectively, 44.34% and 66.12% higher than the expected, those in the magazine, newspaper, and academic registers are 18.36%, 21.02%, and 66.86% lower than the expected, respectively.

TABLE 3
Most Frequent PVs Distribution Across the Registers and the Results of a One-Way Chi-Square Test and Gries's Dispersion Test

	Spoken	Fiction	Magazine	Newspaper	Academic	Total				
Size (million)	78.82	74.88	80.66	76.33	76.20	386.89	*df*	*Chi-square* χ^2	*p*	*Gries's DP*
Raw frequency of PVs	411,326 (+44.34%)*	449,720 (+66.12%)*	244,270 (−18.36%)*	225,079 (−21.02%)*	94,441 (−66.86)*	1,424,836 (1,339,479)**	4	324,445.88	0.0001	0.214 0.268_{norm}
Frequency PMWs	5,218.55	6,005.88	3,028.39	2,948.76	1,239.38	3,683.74	4	324,445.88	0.0001	0.214 0.268_{norm}

Note. *The percentage the observed frequency deviated from the expected frequency. **Rosengren's adjusted frequency produced by Gries's Dispersions2 test.

It is thus clear from the test results that the PVs are much more common in fiction and spoken English than in magazines, newspapers, and, especially, academic writing. The results support the conclusion of Biber et al. (1999, p. 408) on the issue: "Overall, phrasal verbs are used most commonly in fiction and conversation; they are rare in academic prose. In fiction and conversation, phrasal verbs occur almost 2,000 times per million words." The only difference between the finding of Biber et al. and my finding is the rate of occurrence. Although the PV frequency in fiction and conversation in their study is "almost 2,000 per million words," the rates in the two registers found in this study are almost three times that. The reason for this large difference between their number and mine appears, again, to be the narrower definition of PV used in their study, an issue explained earlier (Footnote 3). One can attribute the difference to this reason quite confidently, because the frequency numbers in my study and in Gardner and Davies' (2007) study are quite comparable, and our two studies used the same definition. In Gardner and Davies' study, the frequency of the top 100 PVs in the BNC is 278,780 or 2,788 PMWs (p. 349), a number that would have been even higher if it had included those of the 50 PVs included in my study. Furthermore, given that the 2,788 PMWs frequency is the average that included the much lower frequencies in the newspaper and academic writing registers, one can certainly expect the numbers in their spoken and fiction registers to go much higher than 2,788 PMWs.

Although the overall cross-register analysis provides information about the PVs' general distribution patterns, it does not offer information about the behavioral patterns of the individual PVs, especially those that actually occur more often in the registers other than in fiction or speech. Such information is very useful for language learning. Therefore, I conducted an analysis of each individual PV's raw observed frequencies across the five registers, using both a one-way chi-square test and Gries' (2008b) Dispersions2 test. Although the chi-square test reveals that every one of the 150 PVs showed a significant difference ($p < 0.001$) in its frequencies among the five registers, the dispersion test shows their DP_{norm} values vary substantially, ranging from 0.045 (in the case of *make up*) to 0.74 (in the case of *look up*).[6] This means that, of the 150 PVs, *make up* is distributed most evenly across the registers, whereas *look up* is distributed most unevenly. Based on their DPs, I divided the PVs into three groups: (1) fairly evenly distributed, with a DP_{norm} below 0.25 (65 of them or 43.33%), (2) not evenly distributed, with a PD_{norm} between 0.25 and 0.499 (68 or 45.33%), and (3) very unevenly distributed, with a PD_{norm} of 0.5 or above (17 or 11.33%). The latter two types combine for 85 (56.67%). Each PV's

[6] DP_{norm}, instead of DP, is opted for because of its ability to show the maximal value.

dispersion classification is shown in the appendix with a superscript number 1, 2, or 3 after the PV.

Classifying PVs by dispersion pattern is very useful for language learning because, as Greis (2008a) showed in his literature review, the distributional range of lexical items has a great impact on second language (L2) learners' processing and learning of them. Items that boast a wider and more even distributional range are processed faster than those that have a narrow one, and hence they should be of higher priority for L2 learners. However, this does not mean one can overlook those unevenly distributed PVs in language learning. In fact, the latter PVs occur mostly in one or two registers, and, as such, they are actually very important for English for specific purposes (ESP) learners, who, because of their specific purpose of study, must focus on the register(s) in which these PVs appear most frequently. For example, *carry out* is an unevenly distributed PV because of its very high frequency in academic writing. For students studying academic English, it should thus be very high on their list of PVs to be learned. Before examining in detail the noticeable distribution patterns of the 85 significantly unevenly distributed PVs, it is important again to note that the PVs reported here are lemmatized and many of them are polysemous, just as most PVs are in general. The distribution of the different meanings of a polysemous PV may vary significantly across registers.

For example, *make up* can mean, among other things, compose or constitute (e.g., "Women *make up* 22 percent of the rural labor force in Nicaragua . . ."); decide, when used in "*make up* one's mind" ("Secretary Powell can *make up* his own mind"); compensate (for) (e.g., "The kids *make up* for their lack in experience with enthusiasm"); and fabricate ("Melanie *made up* that story").[7] I examined the meanings of the first 100 tokens of this PV in COCA's spoken and academic registers. A 2 by 5 Chi-Square test of the meaning distributions (reported in Table 4) yielded a very significant result: $\chi^2(df = 4) = 104.52$, $p < 0.0001$, with a Cramer's V of 0.7229, indicating a clear, significant difference between the semantic distributions of *make up* in the two registers. Although the tokens in spoken English show a fairly even division among the four meanings, the tokens in academic writing mean mostly *compose* (79%). This finding suggests clearly that the cross-register distribution of the different meanings of a PV is also important information. Unfortunately, because of lack of space, this study is unable to offer a close examination of this important issue.

Furthermore, as lemmatized lexical items, the 150 PVs are listed without information regarding their uses in different tenses, for example, *make up* versus *made up* versus *making up*. This latter information is very important for language learners or teachers when deciding which form to

[7] All the examples here and in the following are from COCA.

TABLE 4
Distribution of the Major Meanings of the First 100 Tokens of *Make Up* in the Spoken and Academic Registers

	Compensate	*Compose*	*Decide*	*Fabricate*	Other
Spoken	26	12	27	25	10
Academic	18	79	1	2	0

focus on, because the dominant tenses in which specific PVs are used sometimes differ substantially. For instance, in COCA, although *turn out* is used roughly 50% of the time in the past tense, *go ahead* appears 93% of the time in the present tense. Thus PVs like *turn around* may be good items for instruction in teaching the past tense, whereas PVs like *go ahead* may be best used for teaching the present tense. Again, for lack of space, this study is unable to offer a detailed treatment of the tense distribution of the PVs. Clearly, although the lemmatized list of the 150 PVs is a useful source for learning the most common PVs in general, English learners may still need to seek further semantic or usage information of the PVs when learning these PVs. There are some useful sources they can turn to for help in this regard, including PV dictionaries and online sources like the Wordnet Search (Miller, 2008).

Concerning the distribution patterns of the 85 significantly unevenly or very unevenly distributed PVs, almost all of them appear primarily in fiction and spoken English, the two registers that record the highest overall use of PVs. Sixty of the 85 (70.59%) occur mostly in fiction and 22 (25.88%) in the spoken registers. Only three (3.53%) appear significantly more frequently in the other three registers (two in academic writing and one in newspapers). Because of their rarity, the latter three deserve our attention first. The two PVs that occur mainly in academic writing are *bring about* and *carry out*. *Bring about* is used so predominantly in academic writing that its frequency in academic writing (27.44 PMWs) is many times (varying from 3 to 10 times) more than its frequencies in each of the other four registers. Also worth mentioning here is that *point out*, a fairly evenly distributed PV, registers its highest frequency in academic writing as well. It is important to note that *carry out* and *point out* are also found to be used most frequently in academic writing in the study by Biber et al. (1999). The reason that *bring about* is not on their list seems to be that it does not have at least 40 tokens PMWs in any of the registers, the criterion of inclusion in their study. Biber et al.'s (1999) analysis also shows that *take on, take up,* and *set up* are more common in academic writing than in conversation. However, in this study, only *take up* shows this pattern together with *set out,* likely because of the data in the spoken register of COCA are mostly from TV or radio programs, not from conversations, as was the case for Biber et al.'s spoken corpus data. Obviously, all these PVs

deserve attention in academic writing teaching materials. In addition, fairly evenly distributed PVs that have a substantial frequency in academic writing (e.g., *break down, carry on, follow up, make up, rule out,* and *sum up*) should also be considered.

The only significantly unevenly distributed PV that appears most frequently in newspapers is *pay off*, but there are several in the fairly evenly distributed group that claim their highest frequency in the newspaper register: *grow up, take over, shut down, wind up, turn down, fill out,* and *come off*. Most of these PVs are expressions used to describe business dealings. As such, it is understandable that they often find their way into news. This finding helps illustrate the "field"-specific nature of the use of some PVs, an issue Celce-Murcia and Larsen-Freeman (1999, p. 434) have addressed in some detail. Magazine is the only register in which none of the significantly unevenly distributed PVs is used most frequently. Yet, quite a few PVs (7) in the fairly evenly distributed group each record their highest frequency in this register, including *break down, break up, build up, check out, set up, sum up, stand out,* and *take on*. The fact that these PVs all come from the fairly evenly distributed group can perhaps be explained by the mixed nature of this register. Magazine articles cover a variety of topics, and different magazines have different target audiences, making their contents quite diverse.

The 60 significantly unevenly distributed PVs that occur most often in fiction are a very large group, but a majority of them (over 40) are movement or action expressions, for example, *look a/round, look up, sit down, stand up,* and *walk out*. Because describing human actions constitutes a very large part of fiction, it is truly befitting of fiction to make an extensive use of these action PVs. It is again important to note that some of these PVs are polysemous (e.g., *look up*), but they are used mostly as movement descriptions in fiction. Of the first 100 tokens of *look up* in the fiction register, 98 are about upward vision or head movement, as in the example "When Billy opened his eyes and *looked up*, all he could see out the windows were stars." Of course, not all action PVs appear most frequently in fiction. Some are more common in spoken English. For example, *come down, go in*, among others, are used most frequently in spoken English. In fact, the majority of the most frequent PVs in fiction also show a high degree of frequency in spoken English and vice versa, largely due to the fact that a substantial portion of fiction is made up of dialogues. Still, some action PVs occur almost exclusively in fiction, with just a few tokens (all in the single digits) in the other registers, including *call out, hang up,* and *sit back,* and they are largely mono-meaning, used for depicting actions. In contrast, polysemous PVs that can be used to either describe actions or express other meanings are common in both spoken English and fiction.

Another PV that bears some discussion here is *come on*. In Biber et al.'s (1999) study, it is the most frequent PV in spoken English, but in COCA

and the BNC, *go on* is the most frequent spoken PV. What is particularly striking about *come on* is that its frequency in spoken English in the LSWE corpus (the corpus Biber et al. used) is over 300 PMWs; 266.97 PMWs in the spoken part of BNC; but only 83.67 PMWs in the spoken register of COCA. The most likely explanation for its extremely high frequency in the LSWE and the BNC is that, as pointed out earlier, the data in the spoken register in the two corpora are primarily taken from face-to-face conversations, whereas the data in the spoken register of COCA consist mostly of public speech mediums like radio or TV broadcasting, a much more formal type of spoken language. The corpus examples of *come on* provided by Biber et al. (1999), such as "*Come on*, let Andy do it" and "*Come on*, let's go" help demonstrate the conversational nature of their spoken corpus data.

CONCLUSION: IMPLICATIONS AND LIMITATIONS

This study has offered a comparative examination of the usage patterns of the most frequently used PVs in American and British English and across registers. Besides validating many of the results of Biber et al.'s (1999) and Gardner and Davies' (2007) studies, it has provided some new information about the use of PVs and a comprehensive list of the most common PVs in American and British English, one that complements those offered by Biber et al. and Gardner and Davies with more items and more usage information. In addition, it also presents a cross-register list of the most frequent PVs, showing in which register(s) each of the PVs is used primarily. English learners or teachers can elect to use the lists of the 150 most common PVs in ways that best meet their learning purposes. For example, for a language curriculum or program with a general learning purpose, either the American or the British overall frequency list may be used as a reference guide, depending on whether American or British English is chosen as the target English variety. For ESP programs, however, one of the register-specific frequency lists (e.g., newspapers or academic texts) can be used as the guide. Currently, the frequency order is based entirely on the PVs' overall frequency in COCA. To derive the correct frequency order of a register-specific list, one can copy the desired register list together with the PV items, place them in an Excel spreadsheet, and have the rank order adjusted according the PVs' frequencies in the register using the sorting function. The following are some additional pedagogical implications.

1. Although there are some PV usage and frequency differences between American and British English, the most common PVs are generally rather similar between the two English varieties. Thus, except for those aforementioned usage differences, learners or teachers of English need not worry about the problem of learning PVs that are useful only in American or British English.

2. Although PVs that show a wider and more even distribution across registers usually should receive more attention and perhaps be learned first, unevenly distributed PVs may actually deserve special attention for ESP learners, because of their high frequency in the register(s) that are the ESP learners' focus.
3. Learners of English should be made aware that the use of PVs is register and field sensitive so they can approach PVs more effectively and appropriately. For students learning academic writing in English, it is important to know that, although PVs are generally not common in formal writing, there are a few PVs (e.g., *carry out* and *point out*) that are actually very useful in academic writing, and it will be to the students' advantage to gain command of them. Writing teachers may want to purposely include these PVs in their teaching.
4. Learners should focus mostly on polysemous or idiomatic PVs, because mono-meaning and literal meaning PVs are not only easy to understand but are limited in context and function, as shown in the usage patterns of action PVs uncovered in this study.
5. Learners should also understand that the various meanings and functions of polysemous PVs are also often register specific, as in the case of *make up* discussed earlier. Learners or teachers should consult various sources such as PV dictionaries and online sources like Wordnet 3.0 to become familiar with the different meanings, especially the key meanings of the PVs they are learning.
6. Learners can also take advantage of free online corpora such as COCA and the BNC as useful sources for learning and practicing PVs, especially their different meanings. For example, students can enhance their ability in distinguishing the different meanings of a PV by going through concordance lines of a PV query to determine the meaning of each specific token. Such exposure to PVs can also help learners become more familiar with PVs and then more comfortable in using them, hence helping overcome their inclination to avoid PVs. Furthermore, some useful strategies for learning PVs have been suggested, such as studying the cognitive motivation of the use of the AVPs in PVs to help better grasp the meanings of PVs (Kővecses & Szabó, 1996) and examining the typical noun collocates of PVs to better understand and retain idiomatic PVs.

Limitations of the Study and Implications for Future Research

First, limited by space and research design, this study provides only the lemmatized most common PVs, and it does not provide an examination of the use of the various meanings of those polysemous PVs across various registers. A tense-specific list and an analysis of the various meanings of the PVs can help better understand how the various tenses and meanings of a PV are used, including information such as in

which register or registers each of its tense forms and meanings is most frequently used. Second, as is the case in Biber et al. (1999), the cross-register comparative study of PVs in this study covers only broad categories, offering little information on the PVs usage patterns in specific fields, such as air-traffic control. In the future, more field-specific comparative studies are needed.

ACKNOWLEDGMENTS

I thank the three anonymous reviewers and *TESOL Quarterly* Editor Alan Hirvela for their extremely valuable comments and suggestions. They have helped me significantly enhance the quality of this article. I also thank one of the reviewers for suggesting the dispersion/adjusted frequency test and Professor Stefan T. Gries for allowing me to use his Dispersions2 program.

THE AUTHOR

Dilin Liu is Professor and Director of the Applied Linguistics/TESOL program at the University of Alabama. His main research interests include the learning and teaching of lexis and grammar, especially corpus-based description and learning of lexicogrammar.

REFERENCES

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and usage*. Cambridge, England: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London, England: Longman.

Bolinger, D. (1971). *The phrasal verb in English*. Cambridge, England: Cambridge University Press.

Cambridge international dictionary of phrasal verbs. (1997). Cambridge, England: Cambridge University Press.

Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course* (2nd ed.). Boston, MA: Heinle & Heinle.

Cordon, N., & Kelly, P. (2002). Does cognitive linguistics have anything to offer English language learners in their efforts to master phrasal verbs? *ITL Review of Applied Linguistics, 133–134*, 205–231.

Dagut, M., & Laufer, B. (1985). Avoidance of phrasal verbs: A case for contrastive analysis. *Studies in Second Language Acquisition, 7*, 73–79. doi: 10.1017/S0272263100005167.

Darwin, C. M., & Gray, L. S. (1999). Going after the phrasal verb: An alternative approach to classification. *TESOL Quarterly*, *33*, 65–83. doi: 10.2307/3588191.

Davies, M. (2005). BYU–BNC (an online interface with the British National Corpus: World Edition). Provo, UT: Brigham Young University. Retrieved from http://corpus.byu.edu/bnc/x.asp.

Davies, M. (2008). The corpus of contemporary American English. Provo, UT: Brigham Young University. Retrieved from http://www.americancorpus.org/.

Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly, 41*, 339–359.

Gries, S. T. (2008a). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics, 13*, 403–437. doi: 10.1075/ijcl.13.4.02gri.
Gries, S. T. (2008b). Dispersions2 (statistical program) Retrieved from http://www.linguistics.ucsb.edu/faculty/stgries.
Gries, S. T. (2010). Useful statistics for corpus linguistics. In A. Sánchez & M. Almela (Eds.), *A mosaic of corpus linguistics: Selected approaches* (pp. 269–291). Frankfurt am Main, Germany: Peter Lang.
Hulstijn, J. H., & Marchena, E. (1989). Avoidance: Grammatical or semantic causes? *Studies in Second Language Acquisition, 11*, 241–255. doi: 10.1017/S0272263100008123.
Kővecses, Z., & Szabó, P. (1996). Idioms: A view from cognitive linguistics. *Applied Linguistics, 17*, 326–355. doi: 10.1093/applin/17.3.326.
Laufer, B., & Eliasson, S. (1993). What causes avoidance in L2 learning: L1–L2 difference, L1–L2 similarity, or L2 complexity? *Studies in Second Language Acquisition, 15*, 35–48. doi: 10.1017/S0272263100011657.
Liao, Y., & Fukuya, Y. J. (2004). Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning, 54*, 193–226. doi: 10.1111/j.1467-9922.2004.00254.x.
Liu, D. (2003). The most frequently used spoken American English idioms: A corpus study analysis and implications. *TESOL Quarterly, 37*, 671–700. doi: 10.2307/3588217.
Liu, D. (2008). Linking adverbials: An across-register study and its implications. *International Journal of Corpus Linguistics, 13*, 491–518. doi: 10.1075/ijcl.13.4.05liu.
Longman phrasal verbs dictionary. (2000). Harlow, England: Pearson Education
McCarthy, M., & O'Dell, F. (2004). *English phrasal verbs in use*. Cambridge, England: Cambridge University Press.
Miller, G. (2008). Wordnet Online (Version 3.0, first launched in 2006) Retrieved from http://wordnetweb.princeton.edu/perl/webwn.
Moon, R. (1998). *Fixed expressions and idioms in English*. Oxford, England: Clarendon Press.
Oxford phrasal verbs dictionary for English Learners. (2001). Oxford, England: Oxford University Press.
Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London, England: Longman.
Side, R. (1990). Phrasal verbs: Sorting them out. *ELT Journal, 44*, 144–152. doi: 10.1093/elt/44.2.144.
Spears, R. A. (1993). NTC's dictionary of phrasal verbs and other idiomatic phrasal verbs and phrases. Lincolnwood, IL: National Textbook.
Wyss, R. (2003). Putting phrasal verbs into perspective. *TESOL Journal, 12*, 37–38.

APPENDIX

Frequency and Distribution of the Most Frequent Phrasal Verbs in COCA and the BNC Listed According to Their Overall Rank Order in COCA (Frequency Number PMWs)

	Distribution across the registers in COCA					In COCA		In BNC		Rank- order
PVs	Spoken	Fiction	Magazine	Newspaper	Academic	Total	Rank order	Total	Rank order	difference
go on[2]	**316.37**	198.70	96.71	101.65	52.55	153.48	1	148.33	1	0
pick up[2]	108.67	**262.17**	96.13	90.29	23.70	115.40	2	89.95	3	1
come back[2]	**251.75**	163.84	50.15	69.15	11.92	109.44	3	79.91	5	2
come up[2]	**250.51**	102.38	64.42	68.29	18.91	101.48	4	54.97	9	5
go back[2]	**190.98**	154.92	56.15	65.03	19.74	97.31	5	80.27	4	1
find out[2]	**162.08**	99.41	65.88	50.84	22.35	80.43	6	65.88	8	2
come out[2]	**163.22**	90.99	45.35	50.32	11.48	72.51	7	49.99	12	5
go out[2]	**129.91**	109.79	48.47	56.02	9.62	70.77	8	76.52	6	2
point out[1]	79.13	38.72	76.70	62.05	**90.72**	69.71	9	69.51	7	2
grow up[1]**	89.25	61.39	75.95	**97.38**	22.57	69.56	10	18.44	53	43
set up[1]	67.62	55.21	**83.50**	74.48	43.38	65.11	11	103.12	2	9
turn out[1]	**82.87**	64.50	81.89	58.65	33.39	64.58	12	42.64	21	9
get out[2]	**122.05**	107.17	41.84	44.29	6.92	64.43	13	35.28	30	17
come in[2]	**121.23**	99.71	39.30	46.52	10.11	63.36	14	47.91	14	0
take on[1]	71.67	46.99	**71.93**	70.52	48.51	62.17	15	41.79	22	7
give up[1]	65.31	**72.40**	51.38	67.90	23.76	56.11	16	41.66	23	7
make up[1]	54.35	**61.65**	59.20	56.85	46.91	55.80	17	54.43	10	7
end up[1]**	**82.44**	47.34	63.97	58.23	20.38	54.80	18	33.62	32	14
get back[2]	90.61	**92.45**	35.37	44.52	5.33	53.56	19	45.31	19	0
look up[3]	16.85	**202.87**	19.96	12.84	4.31	50.24	20	38.53	26	6
figure out[1]**	**70.46**	62.95	52.39	41.92	12.40	48.17	21	2.73	147	126
sit down[2]	55.75	**126.49**	27.65	23.23	6.34	47.43	22	44.57	20	2
get up[2]	51.79	**126.60**	31.22	24.56	5.09	47.41	23	39.18	25	2
take out[3]	57.45	**87.74**	35.15	33.83	8.28	44.32	24	34.10	31	7
come on[3]	83.67	**108.72**	12.84	10.60	1.86	43.22	25	48.07	13	12
go down[2]	**73.51**	64.34	26.47	27.41	6.43	39.62	26	47.59	15	11
show up[1]**	**52.11**	45.79	43.30	47.22	8.90	39.57	27	7.64	119	92
take off[2]	36.73	**81.66**	32.28	27.54	5.71	36.58	28	21.52	46	18

TABLE
Continued

PVs	Distribution across the registers in COCA					In COCA		In BNC		Rank- order difference
	Spoken	Fiction	Magazine	Newspaper	Academic	Total	Rank order	Total	Rank order	
work out[1]	**49.63**	34.09	43.50	39.57	14.62	36.47	29	46.81	16	13
stand up[2]*	42.76	**91.32**	20.94	18.20	5.70	36.46	30	30.43	34	4
come down[2]	**60.32**	54.37	24.25	27.20	6.82	34.58	31	32.90	33	2
go ahead[3]*	**119.46**	26.34	9.69	9.20	2.66	33.80	32	17.47	56	24
go up[2]	**64.84**	35.39	26.80	32.91	5.55	33.23	33	36.61	29	4
look back[2]	33.96	**71.20**	19.91	17.99	7.99	29.97	34	22.40	42	8
wake up[2]**	34.03	**63.82**	26.42	20.42	3.66	29.54	35	16.07	62	27
carry out[2]	26.48	12.13	19.92	23.83	**62.25**	28.86	36	41.60	24	12
take over[1]	30.01	23.28	26.08	**45.57**	16.21	28.24	37	53.95	11	26
hold up[2]	22.52	**75.11**	20.08	17.33	7.26	28.16	38	16.16	61	23
pull out[2]**	22.82	**77.47**	18.45	16.57	3.37	27.42	39	13.99	73	34
turn a/round[2]	33.08	**64.10**	16.08	20.39	4.33	27.37	40	15.62	64	24
take up[1]	21.42	**40.36**	29.42	22.49	22.56	27.19	41	45.86	18	23
look down[3]	9.01	**97.09**	11.98	6.09	3.20	24.96	42	22.11	43	1
put up[1]	**35.49**	33.16	20.53	28.26	4.99	24.49	43	28.22	36	7
bring back[1]	34.78	**35.56**	20.63	21.21	9.59	24.34	44	21.90	44	0
bring up[2]	**44.72**	34.99	17.83	15.73	8.14	24.31	45	24.95	40	5
look out[3]	14.91	**75.72**	15.01	12.36	3.57	23.97	46	16.33	59	13
bring in[1]	**38.97**	21.18	18.24	25.76	11.71	23.23	47	24.93	41	6
open up[1]**	**32.25**	21.53	23.12	19.11	17.33	22.74	48	20.43	49	1
check out[1]**	25.73	28.62	**33.62**	19.74	3.36	22.35	49	5.73	128	79
move on[1]	**35.97**	25.05	18.78	18.92	6.89	21.18	50	14.12	72	22
put out[2]	**38.96**	27.87	16.22	17.91	4.20	21.07	51	16.52	58	7
look a/round[3]	12.78	**73.79**	9.31	7.57	2.17	20.75	52	14.67	68	16
catch up[1]**	21.96	**31.00**	20.22	20.45	8.45	20.39	53	16.05	63	10
go in[2]	**48.10**	31.12	7.86	12.42	2.34	20.37	54	19.65	51	3
break down[1]	19.91	13.98	**26.49**	17.27	17.55	19.15	55	21.89	45	10
get off[2]	25.67	**33.29**	15.46	17.58	2.47	18.85	56	10.81	90	34
keep up[1]**	17.43	**25.45**	21.99	21.72	7.63	18.85	56	13.38	78	22

TABLE
Continued

PVs	Distribution across the registers in COCA					In COCA		In BNC		Rank- order difference
	Spoken	Fiction	Magazine	Newspaper	Academic	Total	Rank order	Total	Rank order	
put down[2]	17.14	**50.11**	13.55	10.55	3.35	18.75	58	28.60	35	23
reach out[2]**	13.77	**45.17**	14.24	13.73	7.72	18.74	59	9.45	104	45
go off[2]	28.05	**37.63**	12.46	12.12	3.19	18.62	60	20.94	47	13
cut off[1]**	20.49	**30.74**	16.89	14.50	7.65	18.01	61	13.74	74	13
turn back[3]	8.17	**64.53**	8.60	6.50	3.44	17.91	62	13.67	75	13
pull up[3]**	9.20	**57.83**	13.14	8.25	1.93	17.81	63	10.53	92	29
set out[1]	12.74	19.40	**23.68**	14.83	17.54	17.67	64	46.11	17	47
clean up[1]**	20.87	**22.09**	17.88	19.81	7.19	17.58	65	9.26	105	40
shut down[1]**	23.60	12.91	17.38	**24.46**	5.91	16.92	66	4.69	143	77
turn over[1]	20.25	**25.27**	13.53	18.20	5.43	16.50	67	9.70	102	35
slow down[1]**	17.34	**22.14**	21.05	13.98	6.71	16.29	68	11.91	85	17
wind up[1]**	19.97	15.72	19.80	**21.13**	3.10	16.02	69	8.26	111	42
turn up[1]	13.19	**26.28**	19.13	14.66	4.92	15.62	70	26.97	38	32
line up[1]**	14.22	**20.94**	17.82	20.40	4.17	15.51	71	9.96	98	27
take back[2]	23.83	**28.47**	10.59	10.44	4.24	15.47	72	16.20	60	12
lay out[1]**	**20.65**	18.59	16.95	11.49	8.44	15.27	73	2.64	148	75
go over[2]	26.11	**32.96**	7.87	7.49	2.22	15.25	74	9.86	101	27
hang up[3]**	8.36	**52.40**	9.27	6.45	0.92	15.23	75	5.40	133	58
go through[2]	**38.16**	12.35	9.20	13.43	2.51	15.22	76	9.66	103	27
hold on[2]	23.76	**28.98**	11.79	8.46	3.12	15.19	77	9.04	107	30
pay off[2]**	15.67	7.60	22.50	**24.60**	4.46	15.09	78	6.18	125	47
hold out[3]	6.84	**49.57**	8.24	7.60	4.34	15.06	79	15.00	67	12
break up[1]	16.87	17.91	**18.47**	14.76	6.31	14.91	80	12.80	81	1
bring out[1]	**22.53**	18.96	13.30	10.66	7.07	14.53	81	14.18	71	10
pull back[3]**	9.67	**45.94**	9.14	6.88	1.76	14.47	82	7.50	120	38
hang on[1]**	13.41	**30.90**	13.50	11.50	2.80	14.34	83	20.11	50	33
build up[1] **	17.01	8.76	**21.55**	12.16	10.99	14.21	84	37.34	28	56
throw out[2]**	**23.19**	18.55	10.33	15.20	3.36	14.13	85	4.91	140	55
hang out[1]**	15.55	**19.47**	17.07	15.67	2.60	14.10	86	2.74	146	60

TABLE
Continued

PVs	Distribution across the registers in COCA					In COCA		In BNC		Rank- order difference
	Spoken	Fiction	Magazine	Newspaper	Academic	Total	Rank order	Total	Rank order	
put on[2]	18.04	**28.87**	11.93	8.69	3.12	14.08	87	14.21	68	19
get down[2]	22.23	**23.62**	10.31	9.59	1.96	13.53	88	15.31	65	23
come over[2]	16.40	**35.41**	8.57	5.99	1.35	13.43	89	9.99	99	10
move in[1]	13.07	**24.99**	11.48	14.94	2.98	13.43	89	7.86	116	27
start out[1]**	**20.52**	10.83	16.43	13.40	4.03	13.14	91	4.88	141	50
call out[3]**	5.28	**43.99**	7.48	6.04	3.53	13.03	92	3.79	144	52
sit up[3]	5.63	**50.71**	6.35	1.83	0.91	12.83	93	11.53	87	6
turn down[1]	13.69	17.09	12.99	**17.15**	2.63	12.71	94	10.46	94	0
back up[1]**	14.22	**18.54**	12.84	13.06	4.29	12.58	95	9.08	106	11
put back[2]	16.26	**25.91**	9.10	8.86	2.80	12.53	96	13.63	77	19
send out[1]**	**18.08**	14.56	12.43	12.08	4.69	12.40	97	13.67	76	21
get in[2]	19.92	**22.12**	8.60	9.66	1.65	12.36	98	11.22	88	10
blow up[1]**	**21.05**	14.46	10.49	11.84	2.55	12.11	99	7.79	117	18
carry on[1]	11.55	**17.41**	12.09	10.43	8.86	12.04	100	38.51	27	73
set off[1]	9.40	**18.54**	14.39	12.35	4.33	11.79	101	18.60	52	49
keep on[2]**	14.25	**20.01**	12.67	9.33	2.31	11.71	102	8.28	110	8
run out[2]*	14.26	**21.46**	9.21	10.97	2.17	11.57	103	11.89	86	17
make out[3]	6.93	**36.47**	7.67	4.34	2.14	11.35	104	11.00	89	15
shut up[3]*	8.88	**38.37**	5.00	3.89	1.16	11.27	105	14.19	70	35
turn off[2]	10.28	**20.22**	13.63	8.99	3.20	11.25	106	5.91	126	20
bring about[2]	10.72	2.80	8.67	6.51	**27.44**	11.22	107	20.73	48	59
step back[3]**	8.16	**32.32**	7.10	4.97	2.74	10.92	108	3.31	145	37
lay down[2]**	8.97	**24.68**	10.56	5.48	5.12	10.89	109	10.08	97	12
bring down[1]	**15.92**	15.46	9.52	9.21	4.19	10.86	110	10.17	95	15
stand out[1]**	7.79	12.47	**13.92**	10.72	9.32	10.85	111	8.14	113	2
come along[2]	15.01	**17.86**	9.26	9.51	1.81	10.68	112	12.64	82	30
play out[1]**	**15.95**	6.49	9.00	11.69	8.27	10.32	113	2.62	149	36
break out[1]	11.49	**12.23**	10.25	11.57	5.76	10.26	114	9.91	100	14
go a/round[2]	**18.56**	17.11	6.32	6.50	1.93	10.07	115	17.96	55	60

TABLE
Continued

PVs	Distribution across the registers in COCA					In COCA		In BNC		Rank- order difference
	Spoken	Fiction	Magazine	Newspaper	Academic	Total	Rank order	Total	Rank order	
walk ou^{2}**	11.99	**24.84**	5.79	6.52	1.35	10.01	116	8.06	115	1
get through^{2}-	**19.06**	11.66	8.11	7.55	1.92	9.70	117	5.31	134	17
hold back^{2}-	8.65	**18.66**	7.72	7.47	3.58	9.16	118	8.19	112	6
write down^{1}**	9.19	**16.77**	8.91	5.02	5.46	9.04	119	15.05	66	53
move back^{1}-	7.66	**14.85**	7.93	9.76	3.14	8.63	120	5.63	131	11
fill out^{1}**-	8.11	9.28	8.58	**10.86**	5.51	8.46	121	2.55	150	29
sit back^{2}-	7.38	**23.69**	5.50	5.29	0.75	8.43	122	8.30	109	13
rule out^{1}**	**11.86**	2.50	8.22	9.88	8.62	8.25	123	13.05	79	44
move up^{1}-	8.23	**10.44**	8.68	10.42	3.26	8.21	124	4.75	142	18
pick out^{2}-	7.76	**16.31**	8.90	5.63	2.44	8.19	125	8.52	108	17
take down^{2}-	10.78	**16.95**	6.01	5.57	1.82	8.19	126	7.71	118	8
get on^{2}	13.91	**14.76**	5.18	5.50	1.58	8.17	127	26.83	39	88
give back^{1}-	**11.88**	10.71	6.38	7.94	2.93	7.97	128	5.05	138	10
hand over^{2}**	7.64	**17.25**	5.95	6.26	3.01	7.96	129	17.35	57	72
sum up^{1}**	6.17	2.96	**11.63**	8.24	9.62	7.77	130	12.28	84	46
move out^{2}-	10.10	**13.22**	5.99	7.42	2.22	7.76	131	5.70	129	2
come off^{1}-	8.83	8.27	7.90	**12.66**	0.67	7.67	132	5.13	136	4
pass on^{1}**	8.26	**8.80**	8.50	6.56	4.88	7.42	133	12.81	80	53
take in^{2}-	6.39	**15.16**	6.26	5.03	2.74	7.07	134	5.07	137	3
set down^{3}-	1.89	**27.64**	2.95	1.43	1.59	6.95	135	5.02	139	4
sort out^{1}**	**9.00**	7.73	6.91	5.50	4.93	6.82	136	27.36	37	99
follow up^{2}**	**13.40**	4.55	3.16	5.45	7.03	6.73	137	10.12	96	41
come through^{1}-	**10.82**	8.64	5.43	6.84	1.54	6.66	138	5.64	129	9
settle down^{2}**	4.17	**15.63**	6.46	4.91	1.75	6.53	139	10.76	91	48
come a/round^{2}	8.25	**14.05**	4.50	4.93	0.97	6.50	140	12.42	83	57
fill in^{1}**	5.65	**7.92**	6.35	6.34	3.74	5.99	141	18.18	54	87
give out^{1}-	**8.42**	7.77	4.80	5.31	1.77	5.62	142	5.30	135	7
give in^{2}-	4.80	**10.95**	6.11	3.88	2.24	5.58	143	5.76	127	16
go along^{2}-	**9.63**	6.37	4.29	4.49	1.56	5.28	144	7.14	123	21

TABLE
Concluded

PVs	Distribution across the registers in COCA					In COCA		In BNC		Rank- order difference
	Spoken	Fiction	Magazine	Newspaper	Academic	Total	Rank order	Total	Rank order	
break off[2]-	2.96	**11.93**	4.56	2.70	1.93	4.77	145	5.46	132	13
put off[1]-	4.80	**6.61**	5.31	5.20	1.44	4.67	146	7.39	121	25
come about[1]-	**7.82**	2.52	4.09	3.31	5.28	4.63	147	7.38	122	25
close down[1]**	**7.60**	3.58	3.33	3.88	2.17	4.13	148	10.48	93	55
put in[2]-	5.09	**7.22**	3.16	3.54	1.05	4.00	149	8.06	114	35
set about[1]-	1.00	**3.22**	3.20	2.07	2.13	2.32	150	6.42	124	26

Note. Superscripted number (1, 2, or 3) after each PV indicates its distribution pattern across the registers: 1 = fairly evenly distributed; 2 = not evenly distributed; 3 = very unevenly distributed. The bold number in each PV entry is the highest among the five registers. *Indicates PV is one of the 4 PVs from Biber et al.'s list that is not on Gardner and Davies' top 100 PV list. **Indicates PV is one of the 33 PVs this study has identified. The minus sign after a PV indicates its number of tokens PMWs is below 10 in either corpus.