Pointing Out Frequent Phrasal Verbs: A Corpus-Based Analysis

DEE GARDNER AND MARK DAVIES
Brigham Young University
Provo, Utah, United States

Supposing is good, but *finding out* is better.
Mark Twain (DeVoto, 1922)

This study attempts to shed new light on one of the most notoriously challenging aspects of English language instruction—the English phrasal verbs. The highest frequency phrasal verb constructions in the 100-million-word British National Corpus are identified and analyzed. The findings indicate that a small subset of 20 lexical verbs combines with eight adverbial particles (160 combinations) to account for more than one half of the 518,923 phrasal verb occurrences identified in the megacorpus. A more specific analysis indicates that only 25 phrasal verbs account for nearly one third of all phrasal-verb occurrences in the British National Corpus, and 100 phrasal verbs account for more than one half of all such items. Subsequent semantic analyses show that these 100 high-frequency phrasal verb forms have potentially 559 variant-meaning senses. The authors discuss how learners, teachers, and materials developers might utilize the findings of the study to improve instruction of phrasal verbs in English language education.

Many English language teachers have noted the importance of multiword knowledge in developing their learners' nativelike fluency (Moon, 1997; Schmitt, 2004; Wray, 2000, 2002). Idioms (e.g., *kick the bucket*), phrasal verbs (e.g., *chew out*), stock phrases (e.g., *how do you do*), prefabs (e.g., *the point is*), and other multiword structures are crucial to English, and they add a definite richness to the language. However, there has been a general confusion regarding which multiword items to teach and the best ways to include them in language training (Condon & Kelly, 2002; Darwin & Gray, 1999; Nesselhauf, 2003) and language assessment (Read, 2000). Of particular concern is the evidence that multiword items "are not learned well through ordinary language experience" (Coady, 1997, p. 282).

During the past two decades, as many researchers and teachers have begun to shift attention from syntax to vocabulary in second language education (Folse, 2004; Laufer, 1997), interest in multiword vocabulary items has been growing. High-powered computers, robust software, and large electronic collections of actual language (called *corpora*; singular, *corpus*) have enabled researchers to better identify and classify these otherwise elusive structures that permeate English as well as many other languages (Hunston, 2002; Moon, 1997; Read, 2004; Stubbs, 2001). However, whereas corpus linguists have been able to provide many more insights regarding these items, it remains clear that the surface of this complex issue has scarcely been scratched (Read, 2004).

Perhaps the most important area for progress in the discussion of multiword items is English phrasal verbs. The study of phrasal verbs promises to provide valuable insights into what many linguists and applied linguists have begun to recognize as a multiword middle ground between "syntax and lexis" that has important ramifications for second language acquisition (Gass & Selinker, 2001, p. 391). Additionally, phrasal verbs are notoriously difficult for nonnative learners to acquire, a problem exacerbated by the fact that they tend to be very common and highly productive in the English language as a whole (Celce-Murcia & Larsen Freeman, 1999; Darwin & Gray, 1999; Moon, 1997). This dilemma is further complicated by the fact that many nonnative English speakers actually avoid using phrasal verbs altogether, especially those learners at the beginning and intermediate levels of proficiency. (See Liao & Fukuya, 2004, for a review of this topic.) Even learners whose native language actually contains phrasal verbs (e.g., Dutch) often avoid using such forms when communicating in English (Hulstijn & Marchena, 1989).

The purpose of the current study is to establish a logical rationale for narrowing the scope of phrasal verbs in English language training based on frequencies of actual occurrence in a large representative corpus of English—*British National Corpus: World Edition* (BNC; British National Corpus Consortium, 2000). With this aim in mind, we are more concerned with probabilities of encounter than with possibilities of acquisition. The latter will be left to the fruitful discussions of how multiword items are stored in and retrieved from the mental lexicon (e.g., Wray, 2002) and to actual testing of such items with learners of English (e.g., Read, 2000). We are also more concerned with establishing a basis for what to teach, rather than how to teach—a topic better suited to accounts with that particular focus (e.g., Condon & Kelly, 2002; Cornell, 1985; McCarthy & O'Dell, 2004; Side, 1990; Wyss, 2003).

Because our study is intended to be purely exploratory and informative in nature, no issues of causality will be tested or implied. However, it is our hope that the rich data resulting from the study can be used in

productive ways to inform English language teaching, materials development, and testing and to provide important information for future empirical studies involving language learners and their actual acquisition of English phrasal verbs.

DEFINITION OF ENGLISH PHRASAL VERBS

It is rare to read an article about phrasal verbs without some discussion of definitions. Historically, linguists have focused much of their attention on characterizing and classifying phrasal verbs based on syntactic and semantic considerations such as single-word replacement, separability, literal versus figurative meanings, and so forth (e.g., Bolinger, 1971). Some attention has also been given to distinguishing between phrasal verbs, prepositional verbs, phrasal prepositional verbs, and free combinations (e.g., Biber, Johansson, Leech, Conrad, & Finegan, 1999).

However, as Darwin and Gray (1999) point out, many of the traditional tests for phrasal verbs have notable exceptions, and experts often disagree on which items to include under this fuzzy grammatical category. This "confusion among sources," according to the researchers, "leads to confusion for students and teachers" (p. 67). Interestingly, however, even their own "alternative approach" to phrasal verb classification has come under criticism on the grounds that their system excludes forms that might "actually aid in the teaching" of phrasal verbs (Sawyer, 2000, p. 151), and that their proposed classification system for phrasal verbs may not do much to advance second language curriculum development and instruction beyond what is already available (Sheen, 2000).

Perhaps missing in this ongoing debate among linguists is the ecological reality of phrasal verb forms in the actual language experience of nonnative speakers of English. In other words, if even the linguists and grammarians struggle with nuances of phrasal verb definitions, of what instructional value could such distinctions be for the average second language learner? Therefore, for purposes of the current study, we will rely on a more functional and objective definition of *phrasal verb* that is stated as follows: all two-part verbs in the BNC consisting of a lexical verb (LV) proper (tagged as VV in the BNC) 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). No other criteria for classification will be used in this study, and we will rely on the good judgment of the readers to decide whether, and for what purpose, the data may warrant further scrutiny.

In terms of examples, our definition would encompass verb units with varying degrees of semantic transparency, including literal (e.g., *sit down;*

sit yourself down), figurative (e.g., *chew out; chew the team out*), and all degrees in between. In other words, we make no distinctions in our analysis based on the semantic transparency of the phrasal verbs. Whereas some evidence suggests that second language learners struggle more with figurative phrasal verbs than literal phrasal verbs (Liao & Fukuya, 2004), it is not our task in the current study to classify or distinguish phrasal verbs in any way; rather, our aim is simply to determine which phrasal verbs have the greatest impact in the language as a whole.

We also recognize, at the outset, that potential limitations are inherent in our definition of phrasal verbs. First, we are relying heavily on the BNC tagging software to accurately identify the AVP functions (e.g., *You can turn in for the night*) as opposed to the prepositional functions that many of these forms also have (e.g., *The police told me to turn in the opposite direction*). However, whereas some tagging inaccuracies may exist, we also assume that these will be very few in number, as the creators of the tagged BNC report a classification error rate of 1.58% (less than 2 in 100) for AVPs and 0.59% (less than 1 in 100) for prepositions (Leech & Smith, 2000). In addition, we acknowledge that any computerized analysis of linguistic forms tends to underestimate the number of meanings elicited by those forms in real-language contexts (cf. Widdowson, 2000).

CORPUS-BASED ANALYSES OF PHRASAL VERBS

Despite the potential limitations of the definition noted earlier, there are strong reasons for pursuing a rationale for dealing with phrasal verbs that utilizes actual frequency of occurrences in the language, rather than more traditional approaches that have often relied on isolated linguistic examples, teachers' intuitions, or random groups of phrasal verbs in language training curricula—a point clearly articulated by Darwin and Gray (1999):

> Another reason for the somewhat arbitrary presentation of phrasal verbs is that very little has been done to determine frequency of particular phrasal verbs. Thus, instructors, curriculum designers, and researchers are left working with what they determine by intuition to be the most common or most needed phrasal verbs. Their intuition, though, may not be correct. (p. 67; cf. Darwin & Gray, 2000)

As stated previously, one of the growing strengths of corpus linguistics is in the identification and classification of multiword units:

> A major part of the patterning revealed by concordances is the extent of phraseology, which is not obvious to speakers, and has indeed been ignored by many linguists. The patterns have been discovered, but not created, by the computer. (Stubbs, 2001, p. 153; cf. Hunston, 2002)

To date, there have been several attempts to identify and classify English phrasal verbs based on corpus findings. Chief among these are the *Longman Dictionary of Phrasal Verbs* (Courtney, 1983), the *Collins Cobuild Dictionary of Phrasal Verbs* (Sinclair & Moon, 1989), *NTC's Dictionary of Phrasal Verbs and Other Idiomatic Verbal Phrases* (Spears, 1993), and the *Cambridge International Dictionary of Phrasal Verbs* (Walter & Pye, 1997). Each of these extensive works attempts to identify a large number of English phrasal verbs and to provide a definition and contextualized example of each form. Implied in all the methodologies is that the compilers have attempted to identify and define important phrasal verbs based on corpus data. Beyond this implication, however, there is very little information about actual frequency data that could be used by teachers, materials developers, and test designers who may wish to prioritize treatment of phrasal verbs based on their relative impact in the language as a whole.

Perhaps the best treatment of relative frequency is contained in Biber et al., (1999), the *Longman Grammar,* where a short subsection of the text is devoted to describing the most prolific verbs that combine with AVPs to form phrasal verbs within seven semantic domains (activity intransitive, activity transitive, mental transitive, etc.). The authors also consider how the relative frequencies of the most prolific phrasal verbs vary by register (i.e., in conversation, fiction, news, and academic texts). They provide similar, but separate, information regarding prepositional verbs. Their criteria for inclusion is that an item must "occur over 40 times per million words in at least one register" of their particular corpus (p. 410). Beyond this, however, they provide no information to allow comparison between their narrow set of 31 phrasal verbs and the next most prolific set, for example, or between their set and other phrasal verbs in general. We do not view this as an oversight, but as function of space constraints and the broader focus and purposes of the *Longman Grammar* text in general. However, we feel that more focused and expanded data analyses of phrasal verbs are warranted and that such analyses can benefit English language teaching, materials development, and assessment. The remainder of our study is dedicated to addressing this issue. The primary aims of our study are

1. To determine the most frequent AVPs.
2. To determine how often these forms function as AVPs versus prepositions.
3. To determine the most frequent lexical verbs used in phrasal verb constructions.
4. To determine how often such verbs function as stand-alone verbs versus phrasal verbs.

5. To determine the extent to which these verbs interact with various AVPs.
6. To establish a list of the most frequent verb-plus-particle combinations based on overall frequency and coverage.
7. To determine the approximate number of word senses associated with each of the most frequent verb-plus-particle combinations.

METHOD

The Corpus

The source of phrasal verbs analyzed in this study is the BNC, described as follows:

> The British National Corpus [BNC] is a one hundred million word snapshot of British English, both spoken and written, at the end of the twentieth century. . . . The corpus contains about 4000 samples from the widest possible range of linguistic productions, automatically annotated with part-of-speech tags by the CLAWS system, and SGML-encoded according to the Text Encoding Initiative's Guidelines. (CD cover)

Data Gathering

The grammatically tagged version of the BNC was first converted into a format that could be imported into Microsoft SQL Server, a powerful relational database program. In short, this process included dividing the corpus into every possible two-, three-, four-, five-, six-, and seven-word chunk, with their accompanying grammatical tags. Table 1 shows an example of how the database appeared after the transformations took place.

The second step in gathering the data consisted of software queries to identify and report every instance where a lexical verb (LV) was followed by an adverbial particle (AVP). The particle could be immediately adjacent to the verb (LV + AVP), within two words (LV + X +AVP), within three words (LV + X + X + AVP), within four words (LV + X + X + X + AVP), within five words (LV + X + X + X + X + AVP), or within six words (LV + X + X + X + X + X + AVP). For example, *taken* (LV) *over* (AVP) on the first data line of Table 1 is a LV + AVP phrasal verb because there are no words between the LV (*taken*) and the AVP (*over*), whereas *took* (LV) *a* (AT0) *step* (NNI) *back* (AVP) on the fifth data line is a LV + X + X + AVP phrasal verb because there are two intervening words (the article *a* and the noun *step*) between the LV (*took*) and the AVP (*back*). Our subsequent analyses of all outcomes indicated the presence of many false

TABLE 1
Sample of Four-Word Chunks From BNC With Grammatical Parts of Speech Identified

Chunk	W 1	POS 1	W 2	POS 2	W 3	POS 3	W 4	POS 4
1	taken	LV	over	AVP	by	PRP	the	AT0
2	taken	LV	up	AVP	by	PRP	The	AT0
3	took	LV	on	AVP	a	AT0	new	AJ0
4	taken	LV	up	AVP	in	PRP	the	AT0
5	took	LV	a	AT0	step	NN1	back	AVP
6	take	LV	you	PNP	back	AVP	to	PRP
7	taken	LV	over	AVP	by	PRP	a	AT0
8	taken	LV	up	AVP	with	PRP	the	AT0
9	take	LV	it	PNP	out	AVP	on	PRP
10	takes	LV	us	PNP	back	AVP	to	PRP
11	taken	LV	up	AVP	by	PRP	a	AT0
12	taken	LV	back	AVP	to	PRP	the	AT0
13	took	LV	over	AVP	from	PRP	the	AT0
14	taken	LV	over	AVP	from	PRP	the	AT0

Note. W = word. POS = part of speech tag: LV = lexical verb (tagged as VV in BNC); AVP = adverbial phrase, AT0 = article, PNP = pronoun, PRP = preposition, NN1 = singular noun.

phrasal verbs beyond the four-word scenario. This fact, coupled with the relatively infrequent occurrences of phrasal verbs with longer separations (mostly a frequency of 1), caused us to limit our attention to those of the two- (*take on*), three- (*take it on*), and four-word (*took the nodules back*) varieties. We should note, however, that legitimate longer separations do exist in the BNC and may be worth studying in the future (e.g., *send your certificate of motor insurance back*).

In the third step, we lemmatized the outcomes, so that all inflectional forms of the same verb were counted together. For example, the forms *look, looks, looking*, and *looked* were grouped under the lemma *LOOK*; the forms *take, takes, taking, taken, took*, and *taken* under *TAKE*; and the forms *give, gives, giving, gave*, and *given* under *GIVE*. Hereafter, we distinguish lemmas from individual types by using all uppercase letters to indicate lemmas (cf. Stubbs, 2002).

Form-Based Data Analyses

To perform frequency and coverage analyses, we created Excel spreadsheets from the query outcomes explained in the preceding section. This procedure allowed us to form lists consisting of frequency rankings, cumulative frequencies, cumulative percentages, and so forth.

Meaning-Based Data Analyses

In recognition of the fact that phrasal verb forms, like individual word forms, have multiple, context-sensitive meanings, we also used WordNet

(Miller, 2003) to establish a more realistic picture of the benefits of establishing lists of high-frequency phrasal verbs for teaching purposes. WordNet (Miller, 2003) is an electronic lexical database that, among other functions, recognizes distinctions between different senses of the same word forms (Fellbaum, 1998). For instance, WordNet contains 10 different senses for the phrasal verb *put out* (to *cause inconvenience,* to give *considerable effort,* to *smother,* to *anesthetize,* etc.) and 8 different senses for the phrasal verb *work out* (to *happen in a certain way,* to *elaborate,* to *exercise,* to *solve,* etc.).

Our intention with this semantic analysis of the data is to avoid the kinds of oversimplification found in many corpus-based vocabulary studies, namely, that frequency of word forms is discussed as though it were equivalent to frequency of word meanings. We view this as one of the fundamental issues to be addressed as corpus linguists attempt to build bridges to language education.

RESULTS AND DISCUSSION

Frequencies of AVPs

Table 2 shows the total number of all BNC grammatical tags for each of the 16 preposition-particle forms. It also indicates the number of times each of these forms is tagged as an AVP as opposed to a preposition or

TABLE 2
Frequency of Adverbial Particles (AVPs) in BNC

Form	Total tags	# as AVP	% as AVP
out	149,727	145,706	97.3
up	180,792	158,064	87.4
down	91,832	72,709	79.2
back	97,154	75,233	77.4
off	67,479	37,751	55.9
round	30,821	10,895	35.3
along	18,555	4,925	26.5
over	128,304	32,526	25.4
around	43,391	10,384	23.9
on	705,790	54,956	7.8
through	81,184	5,797	7.1
about	190,615	12,587	6.6
in	1,845,077	34,411	1.9
under	60,049	313	0.5
by	504,969	371	0.1
across	24,053	13	0.1
Total	4,219,792	656,641	15.6*

Note. # = token frequency. * = Average of column.

other grammatical structure, such as the noun *out* (in sports language) or the noun *back* (body part). The overall totals indicate that these forms function 15.6% of the time as AVPs. This fact, coupled with the sheer number of particle forms overall (656,641), suggests that phrasal verb (PV) constructions (comprised of verbs plus AVPs) constitute a major grammatical class. To put this in another light, learners will encounter, on average, one in every 150 words of English they are exposed to, or roughly 2 per average page of written text (i.e., assuming 300 words per page). Of course, these exposure estimates are for the corpus as a whole and could vary somewhat based on register type (informal speech, narrative fiction, expositions, etc.)—a point which future research should more carefully consider.

It is also clear from the values in Table 2 that certain forms are more likely to act as particles than as prepositions. In particular, *out* (97.3%), *up* (87.4%), *down* (79.2%), and *back* (77.4%) occur much more often as AVPs in PV constructions (e.g., *she picked out a birthday card*) than they do as prepositions in prepositional phrases (e.g., *she ran out the door*). Learners who understand this concept would have a definite advantage in recognizing PV forms containing these highly prolific AVPs. In contrast, certain forms, such as *under* (0.5%), *by* (0.1%), and *across* (0.1%), will rarely appear as particles in PV constructions. Learners could also put this information to good use.

From a particle approach, the most difficult forms to deal with appear to be words like *on* (7.8% as particles) and *in* (1.9% as particles) that function much more often as prepositions in general but also appear to have substantial frequencies as particles (*on* 54,956; *in* 34,411). This dilemma suggests the need for learners to have other methods for recognizing PVs besides particle identification and leads naturally to a discussion of the verbal component in PV constructions.

Frequencies of Lexical Verbs in Phrasal Verb Constructions

Table 3 provides information relative to all lexical verb (LV) forms tagged in the BNC, as well as those that function in PV constructions. Approximately 5% of all LV tokens function in PV constructions, or one in every 20. Perhaps this finding becomes more meaningful when one considers that PVs, as a grammatical class, have a higher overall frequency than the verb *are*, the determiners *this* or *his*, the negative *not*, the conjunction *but*, or the pronoun *they*. Looking at this from still another angle, there are 10,404,107 tagged LVs in the 100-million-word BNC, or approximately 1 in every 9.6 words, on average. By extrapolation, this means that a learner will encounter 1 PV, on average, in every 192 words (9.6 × 20) of English, or nearly 2 per page of written text.

It is also important to note that only 8% (1,572 lemmas) of the 19,682

TABLE 3
Descriptive Statistics for BNC Lexical Verbs (LVs) Functioning in Phrasal Verbs (PVs)

Measurement	Total #	# in PVs	% of total
LV tokens	10,404,107	518,923*	5.0
LV lemmas	19,682	1,572	8.0
Average lemma frequency	529	330	
PV lemmas	12,508		

Note. # = token frequency. *Lemma* = all inflectional forms of a verb (e.g., *look, looks, looking, looked*) considered to be the same verb (e.g., *LOOK*). *Average lemma frequency* = calculated by dividing number of tokens by number of lemmas. *PV lemma* = a distinct phrasal type comprised of a lexical verb lemma and an (e.g., *LOOK up, LOOK out, LOOK over* = three distinct phrasal verb [PV] lemmas). * = Count based on PVs that are contiguous (verb [V] + adverbial phrase [AVP]), separated by one word (V + X + AVP), and separated by two words (V + X + X + AVP).

total LV lemmas function in PV constructions. Whereas 1,572 is still a staggering figure from a teaching and learning perspective, it is certainly more manageable than trying to figure out which of the 19,683 different LV lemmas function in PV constructions. Perhaps of more pedagogical concern, however, is the fact that, on average, any given LV lemma functioning in PV constructions repeats only 330 times in 100 million words, or once in every 303,030 words. Also problematic, from a teaching and learning perspective, is that at least 12,508 distinct PV lemmas exist in the BNC alone. For instance, the LV lemma *PUT* combines with at least 15 different particles in the BNC to form PV lemmas (*PUT out, PUT up, PUT on,* etc.). It is therefore crucial to establish a frequency ranking of PVs to determine if some are noticeably more prolific than others.

Verb-Particle Frequencies Involving Top 20 Lexical Verbs

Table 4 displays a frequency ranking of the top 20 LVs found in PV constructions. Perhaps the most important statistic in the table is that these 20 verbs are found in 53.7% of all PVs in the BNC; in other words, more than half of all the PVs contain a verb from this short list. Additionally, the PVs containing these 20 LVs make up 2.7% of all the LVs that learners would encounter in the BNC (279,882 ÷ 10,404,717)—approximately 1 in every 39. Interestingly, several of the verbs in the table appear more often in PV constructions than as stand-alone LVs (i.e., *PICK,* 70.0%; *POINT,* 52.0%; *CARRY,* 51.1%).

It is also noteworthy that many of the LVs in the list are among the most prolific in the BNC, with six of the verbs ranked in the top 10 (*GO,* 2; *GET,* 3; *MAKE,* 4; *TAKE,* 7; *COME,* 9; *GIVE,* 10), and four others in the top 20 (*LOOK,* 11; *FIND,* 13; *PUT,* 16; *WORK,* 19). The fact that these 20 prolific verbs function 24.2% of the time in PV constructions underscores the importance of English PVs, in general, and establishes the

TABLE 4
Descriptive Statistics of Top 20 Lexical Verb (LV) Lemmas Functioning in Phrasal Verb (PV) Forms

LV lemma	# in BNC PVs	% of all BNC PVs	Cum % of all BNC PVs	Total # in BNC	BNC rank	% as PVs
GO	48,016	9.3	9.3	227,103	2	21.1
COME	36,878	7.1	16.4	145,047	9	25.4
TAKE	22,970	4.4	20.8	173,996	7	13.2
GET	20,223	3.9	24.7	213,726	3	9.5
SET	18,569	3.6	28.3	39,149	40	47.4
CARRY	15,617	3.0	31.3	30,572	53	51.1
TURN	13,040	2.5	33.8	44,051	32	29.6
BRING	12,514	2.4	36.2	42,567	33	29.4
LOOK	12,226	2.4	38.6	109,110	11	11.2
PUT	11,970	2.3	40.9	67,839	16	17.6
PICK	9,997	1.9	42.8	14,274	138	70.0
MAKE	7,368	1.4	44.2	210,880	4	3.5
POINT	7,159	1.4	45.6	13,767	149	52.0
SIT	7,112	1.4	47.0	27,388	64	26.0
FIND	6,934	1.3	48.3	96,010	13	7.2
GIVE	6,174	1.2	49.5	125,312	10	4.9
WORK	5,985	1.2	50.6	63,104	19	9.5
BREAK	5,428	1.0	51.7	18,642	109	29.1
HOLD	5,403	1.0	52.7	46,773	30	11.6
MOVE	5,197	1.0	53.7	37,820	41	13.7
Total	278,780	53.7	53.7	1,747,130	39*	24.2*

Note. # = token frequency. Cum % = cumulative frequency percentage. Values based on nonseparable and separable counts (i.e., verb [V] + adverbial phrase [AVP], V + X + AVP, V + X + X + AVP). Total phrasal verb (PV) tokens in the British National Corpus (BNC) = 518,923; total lexical verb (LV) tokens in BNC = 10,404,107. * = Average of column.

value of this particular list for instructional purposes. To further illustrate this point, we must first reiterate that LV lemmas functioning in PV constructions repeat only 330 times on average in a corpus of 100 million words (see Table 3). In stark contrast, these 20 high-frequency LV lemmas repeat 13,994 times on average when functioning in PV constructions (279,882 ÷ 20).

Table 5 provides the raw counts of verb-plus-particle constructions for the top 20 LV lemmas. Considering cumulative frequency percentages, it is clear that these 20 LV lemmas combine with only eight particles (*out, up, on, back, down, in, over*, and *off*)—a total of 160 combinations—to account for more than half (50.4%) of the PVs in the BNC. It is equally clear, however, that the actual verb + particle combinations are highly idiosyncratic. For example, the particle *on* combines with the verb lemma *GO* a total of 14,743 times but never once with the lemma *POINT*, and the particle *over* combines often with the verb *TAKE* (5,158) but never with the verbs *SET, POINT*, or *FIND*. Also, several combinations with higher frequencies contain particles that are infrequently used in PVs in general (e.g., *BRING* <u>*about*</u>, 2,083; *GO* <u>*round*</u>, 1,366; *COME* <u>*along*</u>, 1270).

TABLE 5
Verb-Particle Frequencies of Top 20 Lexical Verbs Functioning in Phrasal Verb (PV) Forms (*Continued on p. 351*)

Verb	Out	Up	On	Back	Down	In	Off	Over
GO	7,688	3,678	14,903	8,065	4,781	1,974	2,104	991
COME	5,022	5,523	4,830	8,029	3,305	4,814	518	1,004
TAKE	3,426	4,608	4,199	1,628	775	509	2,163	5,420
GET	3,545	3,936	2,696	4,552	1,538	1,127	1,086	293
SET	4,633	10,360	11	265	504	281	1,869	1
CARRY	10,798	36	3,869	172	84	32	170	131
TURN	4,284	2,710	292	1,373	1,051	149	594	975
BRING	1,425	2,507	390	2,200	1,022	2,505	31	129
LOOK	1,641	3,871	244	2,251	2,221	250	2	207
PUT	1,660	2,835	1,428	1,369	2,873	810	742	76
PICK	856	9,037	35	3	3	1	44	18
MAKE	1,105	5,469	25	270	65	16	277	75
POINT	6,984	104	0	7	56	0	6	2
SIT	191	1,158	118	834	4,478	145	1	3
FIND	6,619	33	9	128	34	57	4	5
GIVE	532	4,186	34	507	11	579	121	198
WORK	4,703	334	411	36	98	182	33	31
BREAK	996	1,286	3	4	2,199	220	549	2
HOLD	1,507	1,624	908	823	369	34	91	40
MOVE	573	477	1,419	566	306	790	242	201
Total	68,188	63,772	35,824	33,082	25,773	14,475	10,647	9,802
% of PV	13.1	12.3	6.9	6.4	5.0	2.8	2.1	1.9
Cum %	13.1	25.4	32.3	38.7	43.7	46.5	48.5	50.4

These phenomena suggest that certain semantic constraints exist in the possible combinations of LVs and AVPs in English. Pedagogically, therefore, English language learners would have to learn more than the 320 verb–particle combinations depicted in Table 5 (i.e., 20 verbs × 16 particles); they would also have to understand which combinations are less likely to occur or do not exist at all. However, given the prolific nature of this set of verbs and particles in general, such an effort is certainly justifiable, especially if one considers the more random nature of typical PV selection for instructional purposes.

Most Frequent Phrasal Verbs in BNC

Appendix A contains the 100 most frequent PVs in the BNC along with their statistical information. This list is simply a reorganization of the highest verb + particle combinations from Table 5. However, this repurposing of the data offers additional insights regarding the impact of these high-frequency phrasal forms. To facilitate this discussion, we refer the reader to Table 6, which is a consolidation of the statistical information found in Appendix A.

First, it is noteworthy that only 25 PV lemmas make up nearly one

TABLE 5
Verb-Particle Frequencies of Top 20 Lexical Verbs Functioning in Phrasal Verb (PV) Forms (*Continued from p. 350*)

Round	About	Through	Around	Along	Under	By	Across	Total
1,366	244	972	394	717	95	44	0	48,016
1,107	741	567	139	1,270	2	7	0	36,878
78	2	31	37	94	0	0	0	22,970
365	102	533	241	163	3	42	1	20,223
0	645	0	0	0	0	0	0	18,569
10	29	127	107	52	0	0	0	15,617
1,146	38	1	423	0	4	0	0	13,040
105	2,083	11	18	88	0	0	0	12,514
694	45	21	779	0	0	0	0	12,226
21	35	90	16	1	9	0	5	11,970
0	0	0	0	0	0	0	0	9,997
16	0	40	2	8	0	0	0	7,368
0	0	0	0	0	0	0	0	7,159
34	18	4	126	1	1	0	0	7,112
3	3	10	29	0	0	0	0	6,934
3	1	0	2	0	0	0	0	6,174
23	0	100	19	5	10	0	0	5,985
0	0	169	0	0	0	0	0	5,428
2	0	0	1	0	4	0	0	5,403
19	178	2	340	84	0	0	0	5,197
4,992	4,164	2,678	2,673	2,483	128	93	6	278,780
1.0	0.8	0.5	0.5	0.5	0.0	0.0	0.0	53.7
51.4	52.2	52.7	53.2	53.7	53.7	53.7	53.7	

Note. Cum % = cumulative frequency percentage.

third (30.4%) of all PV occurrences in the BNC. Fifty PV lemmas constitute 42.7% of the total (see "Cum % Tot PV"), and only 100 are needed to cover more than one half (51.4%) of all PV occurrences in the BNC. To put this observation into a more practical perspective, language learners familiar with this list of 100 distinct PV lemmas would, on average, be able to negotiate more language containing PV constructions

TABLE 6
Frequency and Coverage Summary of Top 100 Phrasal Verb (PV) Lemmas in BNC

PV rank	#	% of total PV	Cum % total PV	% of total LV	Cum % total PV
1–25	157,921	30.4	30.4	1.52	1.52
26–50	63,437	12.3	42.7	0.61	2.13
51–75	29,485	5.6	48.3	0.28	2.41
76–100	15,781	3.1	51.4	0.15	2.56
Subtotal	266,624	51.4	51.4	2.56	2.56
101–12,508	252,299	48.6	100.0	2.44	5.00
Total	518,923	100.0	100.0	5.00	5.00

Note. PV = phrasal verb. # = token frequency. Cum % = cumulative frequency percentage.

than if they knew the remaining 12,408 distinct PV lemmas in the BNC, which cover only 48.6% of all PV occurrences.

Table 6 also shows that these same 100 PV lemmas account for 2.56% of all lexical verbs in the BNC, or roughly 1 in every 40. In practical terms, this result means that, on average, 1 of these top 100 phrasal verbs will occur in every 400 words of English (1/400)—that is, 10 words of every 100 words in the BNC are lexical verbs (1/10)— and 1 lexical verb of every 40 will be from this high-frequency list of PVs (1/40). Put in this light, the effort of teaching these top 100 PVs certainly appears to be justified, especially when one considers the haphazard manner in which most PV instruction takes place in second language instructional settings.

Word Senses of Frequent Phrasal Verbs

Table 7 displays the word-sense frequencies from WordNet (Miller, 2003) for the top 100 PVs in the BNC. These values underscore the need

TABLE 7
Number of WordNet Senses for Top 100 Phrasal Verbs (PVs) in BNC

PV	Senses	PV	Senses	PV	Senses	PV	Senses
Go on	5	carry on	4	put on	9	move in	3
Carry out	2	go up	7	bring out	9	look around	1
Set up	15	get out	7	move on	1	take down	4
Pick up	16	take out	14	turn back	4	put off	5
Go back	4	come down	5	put back	2	come about	1
Come back	5	put down	7	go round**	5	go along	3
Go out	6	put up	8	break up	19	look round***	0
Point out	3	turn up	5	come along	2	set about	3
Find out	4	get on	7	sit up	2	turn off	3
Come up	12	bring up	8	turn round**	3	give in	2
Make up	8	bring in	5	get in	5	move out	2
Take over	8	look back	2	come round**	1	come through	4
Come out	11	look down*	5	make out	10	move back	1
Come on	5	bring back	2	get off	11	break off	5
Come in	5	break down	8	turn down	5	get through	5
Go down	8	take off	9	bring down	6	give out	4
Work out	8	go off	6	come over	1	come off	3
Set out	3	bring about	5	break out	5	take in	17
Take up	13	go in	1	go over	4	give back	1
Get back	4	set off	7	turn over	9	set down	6
Sit down	3	put out	10	go through	5	move up	2
Turn out	12	look out	2	hold on	5	turn around†	0
Take on	5	take back	6	pick out	2		
Give up	12	hold up	7	sit back	2		
Get up	8	get down	7	hold back	5		
Look up	1	hold out	5	put in	7		

Note. Total senses = 559. PV = phrasal verb. *Consulted *Longman Dictionary of Phrasal Verbs* (Courtney, 1993). **WordNet = *around*. ***See *look around*. †See *turn round*.

in corpus studies to address the semantic as well as the formal characteristics of PVs as we attempt to inform language teaching. In short, the same 100 forms that make up roughly half of all PVs in the BNC have expanded to 559 potential meanings, or 5.6 meanings per PV on average. From our perspective, however, this is still a manageable number for language teachers and materials writers to deal with, especially if we consider the alternative of resorting to the more random lists of PVs—with their questionable utility in addition to their own multiplicity of meanings—that are typically used in English language education.

The noted multiplicity of PV senses also confirms the need for language learners to be exposed to these structures in multiple and varied contexts—a task perhaps aided by the use of concordancing software (Cobb, 1997, 1999; Gavioli & Aston, 2001; Sun & Wang, 2003). The findings also suggest the value of constructing more semantically tagged corpora (Landes, Leacock, & Tengi, 1998) that would allow semantic frequencies to be established for instructional purposes. For instance, the list-high 19 senses of the PV *break up* (see Table 7) could be arranged from highest to lowest semantic frequency, thus prioritizing them for language learning. We acknowledge, however, that corpora of this nature are much easier talked about than constructed.

CONCLUSION AND RECOMMENDATIONS

The primary purpose of this corpus-based study of PVs is to contribute to what Read (2004) describes, in talking about multiword research, as "fresh insights for vocabulary learning and for language teaching" that "may yet transform our understanding of vocabulary and the way it is taught" (p. 156).

We hope that our high-frequency lists of PVs will partially answer the where-do-we-start question so often asked by English language learners, teachers, curriculum designers, and materials developers. We also hope that the findings will make a useful contribution to the growing research base involving multiword language.

Suggestions for Pedagogical Applications

Whereas our concern in this article is not with the actual teaching or acquisition of PVs, we offer the following suggestions regarding how the data of our study might best be used for pedagogical purposes.

1. The concept of *word lemma* is critical to taking full advantage of our lists. Therefore, learners would need to be aware of inflectional re-

lationships in word families (e.g., *TURN down* includes *turn down, turns down, turning down*, and *turned down*).

2. Learners should commit to memory those AVPs that occur much more often in phrasal verb constructions than in prepositional phrases (i.e., *out, up, down*, and *back*) and be taught to look for corresponding verbs, keeping in mind that some combinations can be separated by one or more intervening words (e.g., *turn down the offer* versus *turn the offer down*). Knowing these prolific particles will also allow learners to identify many PVs that occur less frequently in the language in general.
3. Learners should commit to memory our list of the top 20 lexical verb lemmas functioning in PVs that cover 53.7% of all phrasal verb tokens in the BNC (see Table 4). Learners should also know the 16 s indicated in the study and have ample practice in flexibly combining these particles with the top 20 lexical verb lemmas (see Table 5). A special emphasis should be given to the eight most prolific particles (*out, up, on, back, down, in, off, over*) that combine with the 20 lexical-verb lemmas to account for approximately one-half of all phrasal verb tokens in the BNC.
4. Learners should have ample exposure (contextualized and decontextualized) to the top 100 phrasal verb lemmas (see Appendix A), with priority given to the top 25 lemmas (covering nearly one third of all phrasal verb constructions).
5. Learners should be made aware of, and have ample exposure to, the multiple meaning senses that are characteristic of high-frequency PVs (see Table 7). Electronic resources such as WordNet (Miller, 2003) and *VIEW* (Davies, 2005) could be used to ascertain these senses, and provide example contexts for exposure and practice.

Suggestions for Future Research

We also offer the following suggestions for future research based on our experiences in this study.

1. The list of high-frequency PVs in this study must be tested against other megacorpora, as well as more specialized corpora, to establish their validity.
2. A reanalysis of the lists across major registers (e.g., spoken versus written English) and within subregisters of those major groupings (e.g., fiction versus news article reports versus academic prose) could provide additional insights relative to English for specific purposes,

English for academic purposes, content-based instruction, literature-based curricula, and other such pedagogical orientations.

3. Corpus-based research must continue to find better ways of locating, tagging, and counting multiword items. Such measures will surely move us closer to the psychological reality of linguistic forms and the preservation of their meanings. We feel that too many frequency-based vocabulary studies have ignored this concept.
4. Corpus-based research should continue to explore ways of identifying, tagging, and preserving the meaning senses of multiword items, as well as single-word items (cf. Landes, Leacock, & Tengi, 1998). Only then will we be able to more accurately describe natural language in terms of the intricate relationships between linguistic forms and their context-sensitive meanings.
5. Future research regarding frequency of PVs and other multiword items might also benefit from probabilistic analyses such those made possible through Bayesian and frequentist statistical applications (Bod, Hay, & Jannedy, 2003; Manning & Schütze, 1999).

Finally, we recognize that the narrow scope of this study leaves many questions unanswered regarding PVs: What about the large body of less frequent PVs? What about literal versus figurative meanings? What about separable versus nonseparable PVs? What about register variation among PVs? In addition to these and other linguistic questions are the lingering concerns of how to use our growing corpus-based understanding of PVs and other multiword items to affect successful pedagogical outcomes. Recently, several works by noted experts in the field have begun to bridge the gap between corpus-based findings and fruitful instructional practices (e.g., McCarthy & O'Dell, 2004; Schmitt, 2004; Sinclair, 2004). We are certain that many more efforts like these will be needed as we obtain new insights regarding multiword items.

THE AUTHORS

Dee Gardner is an associate professor of applied linguistics and TESOL at Brigham Young University in Provo, Utah, United States. He specializes in vocabulary acquisition, literacy development, and applied corpus linguistics. His current research interests include corpus semantics, including semantic frequency analysis, and strengthening the ties between corpus-based research and English language teaching.

Mark Davies is a professor of corpus linguistics at Brigham Young University, Provo, Utah, United States. He has published widely on corpus design, contruction, and use, as well as variation in syntax. He is the creator of the interface for the 100-million-word British National Corpus and is currently finishing a 360+-million-word, web-accessible corpus of American English, 1990–present.

REFERENCES

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

Bod, R., Hay, J., & Jannedy, S. (Eds.). (2003). *Probabilistic linguistics.* Cambridge, MA: The MIT Press.

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

British National Corpus Consortium. (2000). *British national corpus: World edition.* Oxford: Humanities Computing Unit of Oxford University.

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

Coady, J. (1997). L2 vocabulary acquisition: A synthesis of the research. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition (*pp. 272–290). Cambridge: Cambridge University Press.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System, 25*, 301–315.

Cobb, T. (1999). Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning, 12*, 345–360.

Condon, 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.

Cornell, A. (1985). Realistic goals in teaching and learning phrasal verbs. *International Review of Applied Linguistics in Language Teaching, 23*, 269–280.

Courtney, R. (1983). *Longman dictionary of phrasal verbs.* Harlow, England: Longman.

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

Darwin, C. M., & Gray, L. S. (2000). Comments on Clayton M. Darwin and Loretta S. Gray's "Going after the phrasal verb: An alternative approach to classification": The authors respond. *TESOL Quarterly, 34*, 165–173.

Davies, M. (2005). *VIEW: Variations in English words and phrases.* (An online interface with the *British National Corpus: World Edition*). Provo, UT: Brigham Young University. Available from http://view.byu.edu/

Devoto, B. (Ed.). (1922). *Mark Twain in eruption.* New York: Harper & Bros.

Fellbaum, C. (Ed.). (1998). *WordNet: An electronic lexical database.* Cambridge, MA: The MIT Press.

Folse, K. S. (2004). *Vocabulary myths: Applying second language research to classroom teaching.* Ann Arbor, MI: University of Michigan Press.

Gass, S. M., & Selinker, L. (2001). *Second language acquisition: An introductory course* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Gavioli, L., & Aston, G. (2001). Enriching reality: Language corpora in language pedagogy. *ELT Journal, 55*, 238–246.

Hulstijn, J., & Marchena, E. (1989). Avoidance: Grammatical or semantic causes? *Studies in Second Language Acquisition, 11,* 241–255.

Hunston, S. (2002). *Corpora in applied linguistics.* Cambridge: Cambridge University Press.

Landes, S., Leacock, C., & Tengi, R. I. (1998). Building semantic concordances. In C. Fellbaum (Ed.), *WordNet: An electronic lexical database* (pp. 199–216). Cambridge, MA: The MIT Press.

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T.

Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 20–34). Cambridge: Cambridge University Press.

Leech, G., & Smith, N. (2000). *Manual to accompany the British National Corpus (Version 2) with improved word-class tagging*. Lancaster, England: Lancaster University.

Liao, Y., & Fukuya, Y. J. (2004). Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning, 54*, 193–226.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: The MIT Press.

McCarthy, M., & O'Dell, F. (2004). *English phrasal verbs in use*. Cambridge: Cambridge University Press.

Miller, G. (2003). WordNet (Version 2.0) [software]. Available from http://www.cogsci.princeton.edu/~wn/obtain.shtml

Moon, R. (1997). Vocabulary connections: Multi-word items in English. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 40–63). Cambridge: Cambridge University Press.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics, 24*, 223–242.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics, 24,* 146–161.

Sawyer, J. H. (2000). Comments on Clayton M. Darwin and Loretta S. Gray's "Going after the phrasal verb: An alternative approach to classification": A reader reacts. *TESOL Quarterly, 34*, 151–159.

Schmitt, N. (Ed.). (2004). *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins.

Sheen, R. (2000). Comments on Clayton M. Darwin and Loretta S. Gray's "Going after the phrasal verb: An alternative approach to classification": Another reader reacts. *TESOL Quarterly, 34*, 160–164.

Side, R. (1990). Phrasal verbs: Sorting them out. *ELT Journal, 44*, 144–152.

Sinclair, J. M. (Ed.). (2004). *How to use corpora in language teaching*. Amsterdam: John Benjamins.

Sinclair, J. M., & Moon, R. (Eds.). (1989). *Collins cobuild dictionary of phrasal verbs*. London: Collins.

Spears, R. A. (1993). *NTC's dictionary of phrasal verbs and other idiomatic verbal phrases*. Lincolnwood, IL: National Textbook.

Stubbs, M. (2001). Texts, corpora, and problems of interpretation: A response to Widdowson. *Applied Linguistics, 22*, 149–172.

Stubbs, M. (2002). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.

Sun, Y., & Wang, L. (2003). Concordances in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer-Assisted Language Learning, 16*, 83–94.

Walter, E., & Pye, G. (Eds.). (1997). *Cambridge international dictionary of phrasal verbs*. Cambridge: Cambridge University Press.

Widdowson, H. G. (2000). On the limitations of linguistics applied. *Applied Linguistics, 21*, 3–25.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics, 21*, 463–489.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wyss, R. (2003). Putting phrasal verbs into perspective. *TESOL Journal, 12*, 37–38.

APPENDIX A

Frequency and Coverage of Top 100 Phrasal Verb Lemmas in BNC

Rank	Verb	AVP	#	% of PV	Cum % of PV	% of LV	Cum % of LV
1	GO	on	14,903	2.9	2.9	0.14	0.14
2	CARRY	out	10,798	2.1	5.0	0.10	0.25
3	SET	up	10,360	2.0	6.9	0.10	0.35
4	PICK	up	9,037	1.7	8.7	0.09	0.43
5	GO	back	8,065	1.6	10.2	0.08	0.51
6	COME	back	8,029	1.5	11.8	0.08	0.59
7	GO	out	7,688	1.5	13.3	0.07	0.66
8	POINT	out	6,984	1.3	14.6	0.07	0.73
9	FIND	out	6,619	1.3	15.9	0.06	0.79
10	COME	up	5,523	1.1	17.0	0.05	0.85
11	MAKE	up	5,469	1.1	18.0	0.05	0.90
12	TAKE	over	5,420	1.0	19.1	0.05	0.95
13	COME	out	5,022	1.0	20.0	0.05	1.00
14	COME	on	4,830	0.9	21.0	0.05	1.05
15	COME	in	4,814	0.9	21.9	0.05	1.09
16	GO	down	4,781	0.9	22.8	0.05	1.14
17	WORK	out	4,703	0.9	23.7	0.05	1.18
18	SET	out	4,633	0.9	24.6	0.04	1.23
19	TAKE	up	4,608	0.9	25.5	0.04	1.27
20	GET	back	4,552	0.9	26.4	0.04	1.32
21	SIT	down	4,478	0.9	27.2	0.04	1.36
22	TURN	out	4,284	0.8	28.1	0.04	1.40
23	TAKE	on	4,199	0.8	28.9	0.04	1.44
24	GIVE	up	4,186	0.8	29.7	0.04	1.48
25	GET	up	3,936	0.8	30.4	0.04	1.52
26	LOOK	up	3,871	0.7	31.2	0.04	1.56
27	CARRY	on	3,869	0.7	31.9	0.04	1.59
28	GO	up	3,678	0.7	32.6	0.04	1.63
29	GET	out	3,545	0.7	33.3	0.03	1.66
30	TAKE	out	3,426	0.7	34.0	0.03	1.69
31	COME	down	3,305	0.6	34.6	0.03	1.73
32	PUT	down	2,873	0.6	35.2	0.03	1.75
33	PUT	up	2,835	0.5	35.7	0.03	1.78
34	TURN	up	2,710	0.5	36.2	0.03	1.81
35	GET	on	2,696	0.5	36.8	0.03	1.83
36	BRING	up	2,507	0.5	37.2	0.02	1.86
37	BRING	in	2,505	0.5	37.7	0.02	1.88
38	LOOK	back	2,251	0.4	38.2	0.02	1.90
39	LOOK	down	2,221	0.4	38.6	0.02	1.92
40	BRING	back	2,200	0.4	39.0	0.02	1.95
41	BREAK	down	2,199	0.4	39.4	0.02	1.97
42	TAKE	off	2,163	0.4	39.8	0.02	1.99
43	GO	off	2,104	0.4	40.3	0.02	2.01
44	BRING	about	2,083	0.4	40.7	0.02	2.03
45	GO	in	1,974	0.4	41.0	0.02	2.05
46	SET	off	1,869	0.4	41.4	0.02	2.06
47	PUT	out	1,660	0.3	41.7	0.02	2.08
48	LOOK	out	1,641	0.3	42.0	0.02	2.10
49	TAKE	back	1,628	0.3	42.3	0.02	2.11
50	HOLD	up	1,624	0.3	42.7	0.02	2.13
51	GET	down	1,538	0.3	43.0	0.01	2.14

Frequency and Coverage of Top 100 Phrasal Verb Lemmas in BNC (*Continued*)

Rank	Verb	AVP	#	% of PV	Cum % of PV	% of LV	Cum % of LV
52	HOLD	out	1,507	0.3	43.2	0.01	2.16
53	PUT	on	1,428	0.3	43.5	0.01	2.17
54	BRING	out	1,425	0.3	43.8	0.01	2.18
55	MOVE	on	1,419	0.3	44.1	0.01	2.20
56	TURN	back	1,373	0.3	44.3	0.01	2.21
57	PUT	back	1,369	0.3	44.6	0.01	2.22
58	GO	round	1,366	0.3	44.9	0.01	2.24
59	BREAK	up	1,286	0.2	45.1	0.01	2.25
60	COME	along	1,270	0.2	45.4	0.01	2.26
61	SIT	up	1,158	0.2	45.6	0.01	2.27
62	TURN	round	1,146	0.2	45.8	0.01	2.28
63	GET	in	1,127	0.2	46.0	0.01	2.29
64	COME	round	1,107	0.2	46.2	0.01	2.31
65	MAKE	out	1,105	0.2	46.4	0.01	2.32
66	GET	off	1,086	0.2	46.6	0.01	2.33
67	TURN	down	1,051	0.2	46.9	0.01	2.34
68	BRING	down	1,022	0.2	47.0	0.01	2.35
69	COME	over	1,004	0.2	47.2	0.01	2.36
70	BREAK	out	996	0.2	47.4	0.01	2.37
71	GO	over	991	0.2	47.6	0.01	2.38
72	TURN	over	975	0.2	47.8	0.01	2.38
73	GO	through	972	0.2	48.0	0.01	2.39
74	HOLD	on	908	0.2	48.2	0.01	2.40
75	PICK	out	856	0.2	48.3	0.01	2.41
76	SIT	back	834	0.2	48.5	0.01	2.42
77	HOLD	back	823	0.2	48.7	0.01	2.43
78	PUT	in	810	0.2	48.8	0.01	2.43
79	MOVE	in	790	0.2	49.0	0.01	2.44
80	LOOK	around	779	0.2	49.1	0.01	2.45
81	TAKE	down	775	0.1	49.3	0.01	2.46
82	PUT	off	742	0.1	49.4	0.01	2.46
83	COME	about	741	0.1	49.6	0.01	2.47
84	GO	along	717	0.1	49.7	0.01	2.48
85	LOOK	round	694	0.1	49.8	0.01	2.49
86	SET	about	645	0.1	49.9	0.01	2.49
87	TURN	off	594	0.1	50.1	0.01	2.50
88	GIVE	in	579	0.1	50.2	0.01	2.50
89	MOVE	out	573	0.1	50.3	0.01	2.51
90	COME	through	567	0.1	50.4	0.01	2.51
91	MOVE	back	566	0.1	50.5	0.01	2.52
92	BREAK	off	549	0.1	50.6	0.01	2.52
93	GET	through	533	0.1	50.7	0.01	2.53
94	GIVE	out	532	0.1	50.8	0.01	2.53
95	COME	off	518	0.1	50.9	0.00	2.54
96	TAKE	in	509	0.1	51.0	0.00	2.54
97	GIVE	back	507	0.1	51.1	0.00	2.55
98	SET	down	504	0.1	51.2	0.00	2.55
99	MOVE	up	477	0.1	51.3	0.00	2.56
100	TURN	around	423	0.1	51.4	0.00	2.56

Note. # = token frequency. Cum % = cumulative frequency percentage. AVP = adverbial particle. PV = phrasal verb. LV = lexical verb.