System 76 (2018) 197–209

Contents lists available at ScienceDirect

System

journal homepage: www.elsevier.com/locate/system

Sorting them all out: Exploring the separable phrasal verbs of English

Dee Gardner*, Mark Davies

Department of Linguistics and English Language, Brigham Young University, 4064 JFSB, Provo, UT 84602, USA

ARTICLE INFO

Article history:
Received 8 July 2017
Received in revised form 22 June 2018
Accepted 27 June 2018
Available online 30 June 2018

Keywords:
Separable phrasal verbs
Vocabulary
Corpus
Word lists
Language education
Multiword items

ABSTRACT

Several studies in recent years have addressed the impact of phrasal verbs in the English language in an attempt to improve teaching and inform research. While some experts have mentioned the problematic case of separability with these structures, no major attempt to quantify and qualify separable phrasal verbs (SPVs) has been undertaken. Given the prolific nature of these structures, and the documented difficulty they pose for learning and research, we attempt to partially address this issue by finding, counting, and analyzing the SPVs in a mega-corpus of over 1.3 billion words of American and British English. Specifically, we provide and discuss lists of the most frequent SPVs overall, and for those separated by one, two, and three intervening words. We also provide data regarding separability preferences among prolific SPVs, as well as typical grammatical structures of the intervening words between verbs and their allied particles. Implications for instruction and research are discussed throughout the article.

1. Introduction

One of the significant findings of modern corpus inquiry is that multiword items make up a large portion of the English language as a whole (Conklin & Schmitt, 2008, 2012; Siyanova-Chanturia & Martinez, 2015). Some estimates suggest that they account for as much as 58.6% of spoken English and 52.3% of written English (Erman & Warren, 2000), and that the volume of different multiword forms may actually exceed the number of different individual word forms (Jackendoff, 1995; Mel'čuk, 1995; Pawley & Syder, 1983, pp. 191–225). Compounds, phrasal verbs, idioms, fixed phrases, and prefabs are not only prevalent in the language, but they represent one of the greatest challenges to native-like fluency for nonnative speakers of English (Moon, 1997; Schmitt, 2004; Trebits, 2009; Wray, 2000, 2002).

The case of the English phrasal verb seems to be particularly problematic in this regard (Cornell, 1985; Dagut & Laufer, 1985; Hulstijn & Marchena, 1989; Laufer & Eliasson, 1993; Liao & Fukuya, 2004). For one, there are many different phrasal verbs in the language, some with several figurative meanings, some with several literal meanings, and some with both (Celce-Murcia & Larsen-Freeman, 1999). Adding to this confusion is the fact that phrasal verbs, like many other classes of multiword items, seem to straddle the linguistic fence between "syntax and lexis" (Gass & Selinker, 2001, p. 391). Perhaps nowhere is this more apparent than with the relatively large and seemingly unmanageable class of separable phrasal verbs (SPVs), whose individual components actually become separated in syntax (e.g., *let the dog out*; *let the whole team down*; *let the thief off*).

* Corresponding author.
E-mail addresses: Dee_Gardner@byu.edu (D. Gardner), Mark_Davies@byu.edu (M. Davies).

https://doi.org/10.1016/j.system.2018.06.009
0346-251X/

Ongoing anecdotal accounts from learners and teachers suggest that the syntactic and semantic complexities of such structures pose serious dilemmas for even the most advanced English language learners (cf. Kadia, 1988).

The current study is an attempt to better understand the class of SPVs by determining the most prolific lexical and syntactic forms of such structures in a large corpus of American and British English, consisting of samples from several major registers (spoken, fiction, magazine, newspaper, informal blogs, academic, etc.). We acknowledge from the outset that a truly representative corpus of English, covering all possible registers and varieties of the language, would be ideal. But such a "perfectly representative" corpus does not currently exist anywhere, and such a corpus is probably just a theoretical abstraction (Leech, 2007). But we have used as the basis for this project the very best corpora that are currently available. We believe this to be the largest scale attempt to date to find and count the forms and grammatical compositions of English SPVs—our primary linguistic objective. Our principal applied objective is to provide linguistic resources for English learners, teachers, researchers, and materials developers in language education, while acknowledging that results based on the corpora we have chosen to analyze may be more important to certain learner groups than others (adults more than children; literate individuals more than nonliterate; intermediate and advanced learners more than beginners, etc.).

2. Literature review

2.1. Definition of separable phrasal verbs

As with all studies of this nature, it is crucial to establish a working definition for the language construct under investigation. Historical definitions of phrasal verbs by linguists have focused on semantic and/or syntactic classification issues such as separability, literal versus figurative meanings, single-word replacement, and so forth (e.g., Bolinger, 1971). Corpus linguists have also attempted to distinguish between phrasal verbs, prepositional verbs, phrasal prepositional verbs, and free combinations in their analyses (e.g., Biber, Johansson, Leech, Conrad, & Finegan, 1999).

However, as Darwin and Gray (1999) point out, there is a great deal of confusion and disagreement among experts in terms of which items to include under the fuzzy category of phrasal verbs, and classification often becomes bogged down in ambiguity and inconsistency, eventually leading to frustrations for English language teachers and their students. Ironically, even their own "alternative approach" to phrasal verb classification received some criticism (Sawyer, 2000; Sheen, 2000).

The special case of SPVs is fraught with similar problems of inconsistent and/or incomplete definitions. For instance, traditional grammar explanations of separability or nonseparability often focus on the transitive nature of the verbs and whether a direct object pronoun can separate the verb from the particle in a phrasal verb construction. For example, *looked it up* works, but *came it across* does not (Celce-Murcia & Larsen-Freeman, 1999. p. 428); likewise, *looked it over* works, but *looked it through* does not (Master, 1996, p. 309). However, if we accept that *came across* and *looked through* are phrasal verbs (a highly context- and semantically-sensitive issue), then we must also accept the possibility of their separation by parts of speech other than pronouns (e.g., *came quickly across; came unexpectedly across; looked straight through; looked calmly through*). Furthermore, we must also acknowledge that many transitive phrasal verbs are separated by a pronoun plus some other modifying word (e.g., *sort it completely out; pull it straight out; draw it lightly out*), thus further distancing the particle from its allied verb. If internalized pronouns are taken out of the picture, the separation possibilities can become truly complex, as evidenced by the following actual examples from the British National Corpus (BNC): *send your certificate of motor insurance back*; *ask a couple of other kids out*; *carted all the junior nursing staff off*).

In short, we wish to make a clear distinction in this article between the more traditional issues of separability relating to transitive verbs and intervening direct object pronouns, and our own construct, which is any separation of a lexical verb from its allied adverbial particle. Furthermore, we claim a certain sense of ecological and pedagogical validity with such a definition, as all cases of separability are considered for analysis purposes, as well as for possible pedagogical extensions. From a corpus-linguistic perspective, our definition of SPVs is simply all two-part verbs in our corpus consisting of a lexical verb proper (tagged as VV*) and an adverbial particle (tagged as RP*), where the verb and particle are separated by one, two, or three intervening words. No other criteria will be assumed, including literal versus figurative meanings, transitivity of the verb proper, and so forth.

At the outset, we recognize that our definition relies heavily on the accuracy of the tagging software in identifying adverbial particle (RP*) functions (e.g., *You can't keep a good man down for long*) as opposed to prepositional (I*) functions of the second components of phrasal verb constructions (e.g., *You can't keep a good man down in that environment for long*). However, we assume this to be a relatively minor issue, as the creators of an even earlier version of the CLAWS part-of-speech tagger used in this study report a classification error rate of 1.58% (less than 2 in 100) for adverbial particles, and 0.59% (less than 1 in 100) for prepositions (Leech & Smith, 2000). We also assume that these error rates include instances when an adverbial particle actually functions as a post-nominal modifier, rather than being allied to the verb, as may be the case in the somewhat ambiguous example below.

> 1 US B … soflife.blogspot.com A B C revealed, Hitler died an unhappy man. One may turn into a bully and **harass those around** in order to establish one's superiority over them. [retrieved 3.15.16 from GLOBE US]

Is *around* an adverbial particle in the SPV *harass around*, or is it modifying *those*, as in *those around them*, or in its even more expanded realization, *those who are around them*? In the latter two cases, modern grammatical taggers would tag *around* as a preposition (I*), but in the original case, they may tag *around* as an adverbial particle (RP*), as the CLAWS tagger did in our

study. While a context-by-context analysis of this potential issue is not possible given the size of the corpus, we wish to emphasize that such cases are relatively rare in the data, primarily occurring in low frequency examples like the one above, with a limited set of prepositions/particles like *around, above, below*, and *behind*, and that we must assume they fall within or follow closely the error rates indicated above.

An additional limitation of our form-based definition is that it may create a false sense of equality among SPVs in terms of their potential difficulty for nonnative learners of English. For example, there is good evidence that such learners struggle more with figurative phrasal verbs than with literal phrasal verbs (Liao & Fukuya, 2004), and one of the major criticisms of corpus analysis in general is that it does not always account for such context-sensitive semantic issues (cf. Widdowson, 2000). However, we also believe that a form-based frequency list of SPVs is potentially far more useful in English language education than the haphazard and inconsistent approaches for dealing with such structures that continue to persist, particularly when the anecdotal approach is to "just teach a few interesting ones," or that "English language learners will eventually master the structures through natural exposure to the language." From our viewpoint, the first philosophy is not justified in terms of the specific time-sensitive needs of learners (Darwin & Gray, 1999), and the second is not justified in terms of what we know about the actual acquisition of phrasal verbs (Cornell, 1985; Dagut & Laufer, 1985; Hulstijn & Marchena, 1989; Laufer & Eliasson, 1993; Liao & Fukuya, 2004). Additionally, the potential for frequency of exposure to explain many of the storage, processing, and retrieval issues involved with multiword items in general (Conklin & Schmitt, 2008; Kremmel, Brunfaut, & Alderson, 2017; Siyanova-Chanturia & Martinez, 2015) suggests that a study such as ours is warranted on many practical and theoretical grounds.

2.2. Corpus studies of separable phrasal verbs

Several major publishers have produced dictionaries of phrasal verbs. Examples of these include the Longman Phrasal Verbs Dictionary (2000), the Collins Cobuild Phrasal Verbs Dictionary (2012), NTC's Dictionary of Phrasal Verbs and Other Idiomatic Verbal Phrases (1993), the Cambridge Phrasal Verbs Dictionary (2006), and the Oxford Phrasal Verbs Dictionary for Learners of English (2006). While these large dictionaries are generally corpus-based, include many phrasal verbs, and provide definitions, example sentences, and even some learning activities, there are no specific details regarding rankings and coverage of phrasal verbs that could be used to inform learners, teachers, and materials developers who may be interested in prioritizing which phrasal verbs to teach first, next, and so forth. Perhaps even more applicable to this study is the fact that frequency information regarding SPVs is nonexistent.

Several corpus linguists have attempted to address SPV frequency issues at some level. For instance, Biber, et al. (1999) discuss the issue in terms of "mid-position placement" versus "post-particle placement" of direct objects, concluding that when direct objects are pronouns, they occur "90% of the time" in mid-position (p. 932)—i.e., between the verb and its related particle (e.g., *pick it up* vs. *pick up it*). However, there is no discussion of the actual frequencies of SPVs in their analysis. The two corpus-based studies that actually address frequency issues involving SPVs are Davies (2011) and Liu (2011). In both cases, the researchers identify SPVs for counting purposes, but frequencies of these structures are simply added to the overall totals, with no distinctions or conclusions made for the separable forms. The aim of this study is to address this gap in our understanding of SPVs, with the hope that the information can be used to support English language education and research. Given the many variables associated with SPVs and also those associated with assembling and searching large electronic corpora, we make no claims of cause-effect in this study, and view it as largely exploratory. The following questions will guide the remainder of this study:

1. What are the most frequent SPVs overall, and approximately how many meanings are represented by these frequent SPV forms?
2. What are the most frequent 3-g, 4-g, and 5-g SPVs, and how do they compare? (Note that the term *gram* (more specifically, *n-gram*) is used in computational linguistics to refer to words in a contiguous sequence (chunk) of language. See section 3.2 for more details.)
3. What are the most frequent intervening words in 3-g, 4-g, and 5-g SPVs, and how do they compare?
4. What are the most frequent intervening parts of speech (POS) in 3-g, 4-g, and 5-g SPVs, and how do they compare?

3. Methods

3.1. Description of the corpora

Four grammatically-tagged corpora were used as sources for our analysis of English SPVs.

(Note that most of the data for this study was not obtained via the corpus web interfaces themselves, but via the underlying "full-text" databases, which are available to researchers from http://www.corpusdata.org.)

3.1.1. Corpus of contemporary American English (COCA)

This corpus contains more than 450 million words (tokens) of American English and is equally divided among spoken, fiction, popular magazines, newspapers, and academic texts. It includes 20 million words each year from 1990 to 2012.

(Retrieved from the corpus database on February 4, 2015.) (See also, Davies, 2011.)

3.1.2. British National Corpus (BNC)

This corpus contains approximately one hundred million words (tokens) of British English from the 1980s and early 1990s, consisting of samples from spoken (10%), fiction (16.5%), magazines (7.5%), newspapers (11%), non-academic (17%), academic (16%), and miscellaneous (22%) texts. (Retrieved from the corpus database on February 4, 2015.)

3.1.3. Corpus of global web-based English (GloWbe US)

The entire GloWbe corpus "is composed of 1.9 billion words from 1.8 million web pages in 20 different English-speaking countries" (GloWbE website—Davies, 2013; Davies & Fuchs, 2015). The U.S. portion of GloWbe used in this study consists of approximately 385 million words (tokens) of American English. (Retrieved from the corpus database on February 4, 2015.)

3.1.4. Corpus of global web-based English (GloWbe UK)

The U.K. portion of GloWbe used in this study consists of approximately 385 million words (tokens) of British English. (Retrieved from the corpus database on February 4, 2015.)

The combined corpora used in this study consist of approximately 1.32 billion words (835 million words of American English and 485 million words of British English), covering a range of different informal and formal genres. All corpora were grammatically tagged using the CLAWS system (version 7).

3.2. Data gathering

Four steps were used to prepare the corpus samples for analysis of SPVs. First, the grammatically-tagged corpora were imported into a MS SQL Server, a relational database that serves as the underlying architecture for the corpora. For the purposes of this study, we extracted every three-word, four-word, and five-word string (n-gram) from the corpora, along with the accompanying grammatical tags for each word in each string. Typical 3-word, 4-word, and 5-word strings might look like these examples from COCA:

PICK (VVI) **IT** (PPH1) **UP** (RP)
PICK (VV0) **IT** (PPH1) **ALL** (DB) **UP** (RP)
PICK (VV0) **THE** (AT) **WEDDING** (NN1) **BAND** (NN1) **UP** (RP)

Second, SQL (Structured Query Language) queries were used to report every case, like the ones above, where an adverbial particle (RP) was separated from its lexical verb (VV*) by one word (VV* + X + RP), two words (VV* + X + X + RP), or three words (VV* + X + X + X + RP). Hereafter these structures will be referred to by the total number of grams (words) in the contiguous chunk, so respectively 3-g, 4-g, and 5-g SPVs. Careful visual analysis of many of these structures revealed that the process was largely successful in reporting true SPVs, but there were also cases of false SPVs that could easily be dealt with using certain constraints in the SQL queries (see Appendix A for details). The SPV lists resulting from these adjustments appeared to be very clean based on our visual inspection of many samples, but we acknowledge that some error is unavoidable with this or any machine-based approach used to find and count complex structures in a large electronic corpus. We also note that portions of some 5-g SPVs with conjoined particles were also counted as 3-g SPVs (*turn it on* in *turn it on and off; play it over* in *play it over and over*, etc.). Separations beyond three words (6-g, 7-g, etc.) were not considered in this study because of their low frequency (thus low utility) and the higher degree of error we found in our visual inspections of the machine-generated lists for these longer separations.

The third step in the data preparation involved the lemmatization of the SQL results to ensure that SPV forms containing inflectionally-related verbs were grouped together for counting and analysis purposes. For example, *take it out, takes it out, taking it out, took it out,* and *taken it out* would be counted as being the same phrasal lemma (i.e., *TAKE it out* with a count of five). (Hereafter, all data is reported as *lemmas*, meaning that inflectionally-related types are grouped together for counting, analysis, and reporting purposes.) Finally, all data was imported into Excel spreadsheets for additional sorting and counting.

Assessing the approximate number of meanings (see Question 1 in Section 2.2) was accomplished by entering each of the most frequent SPVs into the online version WordNet 3.1 (Princeton University, 2010), and simply counting the number of meaning senses given. We recognize that some meaning distinctions in WordNet are very minor (Garnier & Schmitt, 2015), but it is beyond the scope of this study to determine which meanings could be conflated, and for which learners.

4. Results and discussion

4.1. Overall frequency of SPV lemmas (research question #1)

Table 1 contains data relative to the overall frequency of SPVs in the 1.32 billion word corpus. Several important statistics are worth noting. First, the search queries identified an astounding 53,390 different phrasal verb lemmas, with 28,556 of

Table 1
Phrasal verb data for overall corpus.

	Tokens	Lemmas	Average Repetition
Phrasal Verbs (All)	8,595,520	52,390	164.1
Separable Phrasal Verbs (with shared)	1,376,463	28,556[a]	48.2
SPVs in 3-Gram Chunks	886,710	20,757	42.7
SPVs in 4-Gram Chunks	381,202	16,133	23.6
SPVs in 5-Gram Chunks	108,551	11,648	9.3
Separable Phrasal Verbs (without shared)		48,538[b]	

[a] This total accounts for shared SPVs within the 3-g, 4-g, and 5-g lists.
[b] This total does not account for shared SPVs within the 3-g, 4-g, and 5-g lists.

these being SPV lemmas (more than half). However, many of the different SPV lemmas occur one time only (11,884), as in the following examples of *purple it up* and *punned it down*.

> 1GB Ghairdyeforum.comABC it looks okay in these pics. I ordered some pink nontheless to try and **purple it up** a bit more, I'll post when I do it. # (retrieved 2.05.17 from GLOBE UK)

> 1US Grc.umd.eduABC heroic syllables both ways. France could not even conquer your great name, But **punned it down** to this facetious phrase – Beating or beaten she will laugh the same (retrieved 2.05.17 from GLOBE US)

The presence of such idiosyncratic usages suggests that the creation of phrasal verbs by some English speakers is alive and well (*gyrate over, cherish up, croon up, convulse along, cuff about*, etc.).

It also provides evidence of the difficulties in determining a starting point for addressing phrasal verbs in language education, and offers additional justification for the frequency-based approach taken in the current study.

Second, approximately 16% (1 in 6) of all phrasal verb tokens (occurrences) are separable, with an average repetition (SPV tokens divided by SPV lemmas) of 48.2 tokens per SPV lemma in the corpus we studied. Furthermore, the nearly 1.4 million SPV tokens means that English users can count on meeting an SPV every ~959 running words, or 1 in every 3.2 pages of written text (i.e., assuming 300 words per page) in the corpora we analyzed.

Third, of the 48,538 total 3–5 g SPV lemmas identified in the corpus, 19,982 (41.2%) are shared between the three lists, leaving 28,556 unique SPV lemmas in the overall corpus. While this number certainly contains many single-occurrence forms, as well as some erroneous forms due to tagging and spelling issues, it is a staggering figure nonetheless.

Finally, there are substantial numbers of all three categories of SPVs analyzed—3-gram (20,757), 4-g (16,133), and 5-g (11,648)—but average repetitions vary greatly from 3-g (42.7) to 4-g (23.6) to 5-g (9.3). This seems logical, given the added complexity of producing longer separations. As Table 2 clearly shows, increased distance between a verb and its allied particle also means increased variability in the intervening slots. Noteworthy here is that a total of 25 intervening words, all pronouns, constitute 99.8% of the intervening possibilities in 3-g SPVs, with only six pronouns (*it, them, him, me, you, her*) accounting for over 85% of all intervening 3-g tokens. Learners and teachers could take advantage of this knowledge when learning and teaching SPVs.

However, coverages of the top 25 intervening phrases for 4-g (11.3%) and 5-g (9.5%) are substantially less, and therefore much less predictable. In fact, the same 99.8% coverage we saw with the 3-g SPVs would require 59,480 different intervening options with 4-g SPVs and 67,878 different intervening options with 5-g SPVs. Clearly, a much different instructional approach is needed to help learners with the more complex SPVs than with the 3-g SPVs. Still, the short lists of 4-g and 5-g intervening words in Table 2 may provide a more justifiable starting point for language training than uniformed (random) selection or presentation. For instance, particular attention could be given to *all the way* (3.4%) in raising learners' awareness of 5-g SPVs, and *it all* (2.4%) in raising their awareness of 4-g SPVs. They might also benefit from studying the interesting case of the 4-g possessive + *way* strings of *his way* (e.g., *made his way back*), *their way* (e.g., *forced their way in*), *my way* (e.g., *know my way around*) and *your way* (e.g., *make your way out*), which, combined, account for 2.0% of the intervening structures in 4-g SPVs. (For more on the "way construction," see Goldberg, 1995 and Israel, 1996).

It is important to note that, in terms of real coverage, the intervening structure *it all* appears in a 4-g SPV once in every 484 pages (on average) based on the corpus we analyzed, and the combined "way" structures appear once in every 581 pages. Individually, these may not seem pedagogically important, but when we consider the alternative of trying to work with tens of thousands of different possibilities with these intervening slots, their utility becomes much more apparent. In fact, a learner is likely to encounter at least one 4-g SPV containing one of the top 25 intervening structures in Table 2 in every 83 pages of text. Why not practice and train with these more prolific structures than with randomly selected examples? We follow this same logic throughout this paper.

4.1.1. Highest frequency SPVs overall (research question #1 cont.)

Table 3 contains the highest frequency SPV lemmas in the corpus when all categories (3-g, 4-g, 5-g) are collapsed into one. Before discussing the data in the table, it is important to note that we calculated impact rankings using two different approaches—a sheer frequency approach (FrgRnk in Table 3), and an average rank approach (AvgRnk in Table 3). The former is perhaps self explanatory, being a simple frequency calculation in the entire corpus, without consideration of possible

Table 2
Top 25 intervening words and phrases in 3-g, 4-g, and 5-g SPVs.

3-Gram Slot 2	% Tot	4-Gram Slot 2 & 3	% Tot	5-Gram Slots 2, 3, & 4	% Tot
it	39.1	it all	2.4	all the way	3.4
them	14.4	the ball	0.9	it over and	0.9
him	10.6	each other	0.8	the whole thing	0.7
me	9.5	his head	0.7	a few year(s)	0.5
you	7.1	his way	0.7	the easy way	0.4
her	5.7	a step	0.7	the other way	0.4
us	3.8	their way	0.6	the same thing	0.3
himself	1.3	them all	0.6	them over and	0.3
themselves	1.0	the word	0.5	a long way	0.2
yourself	1.0	his hand	0.5	the first time	0.2
something	0.9	an eye	0.5	it up and	0.2
myself	0.9	your head	0.4	me up and	0.2
everything	0.7	it back	0.4	a few day(s)	0.2
itself	0.6	my head	0.4	a few month(s)	0.2
one	0.6	the light	0.4	him up and	0.2
herself	0.6	her head	0.4	this over and	0.2
anything	0.4	my way	0.4	her up and	0.2
ourselves	0.4	this one	0.4	it on and	0.1
someone	0.3	its way	0.3	a few thing(s)	0.1
everyone	0.2	a look	0.3	a few week(s)	0.1
nothing	0.2	your way	0.3	that whole thing	0.1
anyone	0.2	the phone	0.3	a few step(s)	0.1
everybody	0.1	that one	0.3	a little far	0.1
somebody	0.1	my hand	0.3	me over and	0.1
anybody	0.1	the f***	0.3	a little bit	0.1
Total %	**99.8**		**11.3**		**9.5**

differences in the four corpora making up the larger corpus (COCA, BNC, GloWbe US, GloWbe UK). The latter takes the average (per 10 million) of each of the four corpora and then averages those averages. In theory this would adjust for those cases where one corpus (e.g., GloWbe UK) might contain a higher proportion of a certain SPV than another corpus (e.g., COCA).

As can be seen in Table 3 (FrqRnk − AvgRnk), there are some differences in the rankings after the top 9 SPV lemmas, with the biggest change being +9 for *put on* (i.e., it would move from rank 19 on FrqRnk to rank 10 on AvgRnk) and +7 for *sort out* (i.e., it would move from rank 30 on FrqRnk to rank 23 on AvgRnk). However, in both cases, the adjustment would only serve to move these two SPVs higher in the list of 31, not remove them from the list altogether. Only the last SPV on the list (*look up*) with a possible adjustment of −1 would be at risk of being removed from the list of 31. Because these differences were so small, we determined to use raw frequency for this and other analyses in the article, but we provide both rankings in the complete listing of SPVs that accompanies this article as an electronic resource (see Appendix A).

Returning to Table 3, the data reveals several key points that inform both research and pedagogy:

1 A total of 31 SPV lemmas of the possible 28,556 cover more than one-fourth (25.3%) of all occurrences (tokens) of SPVs in the corpus. Needless to say, this short list provides a useful starting point to begin dealing with SPV training in language education. Put another way, learners would encounter, on average, at least one of these 31 SPV lemmas in every 12 pages of typical text, based on the corpus we analyzed.
2 These 31 forms have 204 combined meaning senses according to WordNet 3.1 (an average of 6.6 per SPV), with five of the 31 accounting for 67 of the different senses (*pick up*-16, *take out*-14, *set up*-15, *put out*-10, *give up*-12). It is our position that dealing with the 204 different senses to achieve over 25% coverage is still a much better approach than trying to deal with over 28,000 different SPVs, or just picking some interesting ones to study, as is typical in most language instruction and materials development.
3 Despite the fact that these 31 are highly prolific SPVs, they do not all behave the same when it comes to their preference for being separated. For instance, *pick up* is the top SPV lemma in the entire corpus, but it only separates 18.9% of the time, meaning that it occurs much more often as a nonseparable phrasal verb (81.1% of the time) than it does as a separable phrasal verb. It just happens to be a very frequent phrasal verb overall.

The issue of separability preference is shown more clearly in Table 4, which contains a rank order (by percent separable) of the 200 most frequent SPVs. Some on the top of the list are almost always separated, and they even seem odd until we have examples with intervening words—for instance, *make through* (*make it through*), *find back* (*find his way back*), *make down* (*made his way down*), *follow around* (*followed her around*), *want back* (*want her back*), *make back* (*made it back*). Others on the bottom of list like *come up*, *go back*, and *come out* are still in the top 200 SPVs, but they separate much less frequently than they stay together (e.g., *come back up, go all the way back, come on out*). The bolded words in the table also show that overall

Table 3
Highest frequency separable phrasal verbs.

FrqRnk	Verb	AVP	Tot PV	Tot SPV	% Sep	Cum Tot	Cum %	Avg/10 Mil*	AvgRnk**	FrqRnk - AvgRnk	# Senses***
1	pick	up	130,979	24,723	18.9	24,723	1.8	64.56	1	0	16
2	get	out	57,087	20,532	36.0	45,255	3.3	50.51	2	0	7
3	bring	back	35,991	20,495	56.9	65,750	4.8	50.28	3	0	2
4	take	out	43,646	18,951	43.4	84,701	6.2	49.61	4	0	14
5	take	back	24,220	17,989	74.3	102,690	7.5	44.97	5	0	6
6	get	back	72,696	17,869	24.6	120,559	8.8	42.98	6	0	3
7	put	down	26,350	14,977	56.8	135,536	9.8	42.48	7	0	8
8	take	off	40,266	14,481	36.0	150,017	10.9	38.32	8	0	9
9	put	back	15,436	12,701	82.3	162,718	11.8	32.17	9	0	2
10	check	out	52,946	11,065	20.9	173,783	12.6	24.26	14	−4	7
11	set	up	106,217	10,926	10.3	184,709	13.4	25.95	11	0	15
12	figure	out	66,475	10,783	16.2	195,492	14.2	22.55	17	−5	1
13	make	up	85,525	10,579	12.4	206,071	15.0	25.58	12	1	9
14	bring	up	38,060	9577	25.2	215,648	15.7	23.37	15	−1	8
15	help	out	17,709	9425	53.2	225,073	16.4	21.32	18	−3	1
16	get	up	49,324	9386	19.0	234,459	17.0	23.07	16	0	8
17	put	up	37,183	9259	24.9	243,718	17.7	24.84	13	4	9
18	turn	around	26,256	8845	33.7	252,563	18.3	18.86	25	−7	3
19	put	on	21,271	8664	40.7	261,227	19.0	26.27	10	9	9
20	put	out	26,424	8543	32.3	269,770	19.6	21.26	19	1	10
21	keep	up	36,397	8090	22.2	277,860	20.2	18.81	27	−6	5
22	bring	down	16,616	7778	46.8	285,638	20.8	19.97	21	1	6
23	give	up	78,509	7587	9.7	293,225	21.3	18.57	28	−5	12
24	get	down	17,770	7555	42.5	300,780	21.9	20.16	20	4	7
25	let	down	11,475	7267	63.3	308,047	22.4	18.97	24	1	2
26	work	out	60,209	6970	11.6	315,017	22.9	18.83	26	0	8
27	take	down	14,238	6705	47.1	321,722	23.4	17.58	29	−2	4
28	put	in	14,022	6701	47.8	328,423	23.9	19.94	22	6	6
29	turn	off	16,944	6640	39.2	335,063	24.3	16.49	30	−1	3
30	sort	out	18,298	6578	35.9	341,641	24.8	19.65	23	7	3
31	look	up	46,046	6369	13.8	348,010	25.3	15.37	32	−1	1

*Avg/10Mil = avg frequency in all 4 corpora (per 10 million tokens).
**Avg Rnk = rank order based on Avg/10Mil.
***# Senses from WordNet 3.1.

SPV frequency has little to do with separability preference, as these top 10 bolded SPVs are essentially all over in the ranked list.

4.1.2. Highest frequency SPVs by grams (research question #2)

A study of Table 5 suggests that the degree of separability (number of intervening words) might play a role in where some SPVs appear in the rankings. For example, *pick up* accounts for 2.5% of all 3-g SPVs, but only 0.6% of 4-g SPVs, and it does not even appear in the top 50 ranking of 5-g SPVs. As another example, *get out* is the highest ranking 4-g SPV (2.9%) and 5-g SPV (1.9%), but it is 10th on the 3-g list, accounting for only 0.9%. Additionally, there are 16 SPVs in the 3-g top 50 (see bolded words) that are not in the 4- or 5-g lists, nine 4-g SPVs not in the 3- or 5-g top 50, and eleven 5-g SPVs not in the 3- or 4-g top 50. Despite these differences, it is also clear that many of the top 50 SPVs are similar between the 3-, 4-, and 5-g lists. In our view, English training can take advantage of both the common and unique SPVs in the table because of their high utility.

4.2. Grammar of intervening slots in SPVs (research question #3)

As we noted in the discussion of Table 2 (section 4.1), the intervening words in 3-g SPVs are very predictable, with 25 pronouns accounting for 99.8% of all occurrences. We suggested that this low figure could be used to instructional advantage, but that 4-g and 5-g SPVs were much more difficult cases because of the variability possible in the intervening slots of these more complex SPV forms. In fact, we determined that reaching the same 99.8% coverage would require 59,480 different intervening options with 4-g SPVs and 67,878 different intervening options with 5-g SPVs, making a frequency approach much less attractive for teaching about intervening words in SPVs. In this section, we explore the idea of using the frequencies of grammatical parts of speech as a partial solution to this dilemma. Because the 4-g and 5-g SPVs have idiosyncratic tendencies with regards to intervening words between verbs and their allied particles, we will address each one separately. This discussion, as well as others in this paper, may also be informative for corpus and computational approaches that attempt to account for the SPVs of English.

Table 4
Ranking by percent separable (200 highest frequency SPV lemmas).

SPV	% Sep	SPV	% Sep	SPV	% Sep	SPV	% Sep
make through	99.0	call back	47.3	fill in	31.5	shut up	17.7
find back	95.4	suck up	47.3	get over	31.3	move up	17.7
make down	94.9	take down	47.1	cut off	31.1	blow up	17.2
follow around	91.2	set down	46.9	write off	29.8	get through	16.7
want back	90.3	bring down	46.8	turn down	29.3	figure out	16.2
make back	90.1	throw back	46.8	hand over	29.0	keep on	16.0
keep down	87.2	lift up	46.7	work up	29.0	shut down	15.9
ask out	86.2	send down	45.8	pull up	28.8	cut out	15.8
see back	85.7	kick out	45.2	move around	28.6	follow up	15.3
let in	84.3	pull down	44.7	pull over	28.3	throw up	15.1
cheer on	83.8	piss off	44.4	send in	28.1	give out	15.1
keep off	83.0	turn over	43.8	call up	27.2	look over	14.9
put back	**82.3**	**take out**	**43.4**	get in	27.2	run down	14.7
try on	81.9	get down	42.5	track down	26.6	seek out	14.2
see up	81.0	throw down	41.3	bring up	25.2	move back	13.9
laugh off	80.6	push out	41.1	calm down	25.1	look up	13.8
know about	79.6	put on	40.7	tear down	25.0	lay down	13.3
see through	77.6	push up	39.9	put up	24.9	break down	13.1
find in	75.4	drop off	39.8	**get back**	**24.6**	fit in	12.6
take back	**74.3**	turn off	39.2	bail out	24.4	clean up	12.5
put through	73.0	bring on	38.8	get off	24.2	rule out	12.5
think over	72.9	carry around	38.6	wear out	24.1	make up	12.4
leave on	71.7	knock down	38.4	send out	23.7	take up	12.2
keep out	70.2	send off	38.3	leave out	23.4	work out	11.6
work off	69.1	stick out	38.0	freak out	23.4	start off	11.4
see out	67.9	beat up	37.7	clear up	23.3	set off	11.1
top off	66.4	shake up	37.2	pull out	22.7	wake up	10.4
say out	66.3	throw out	36.7	keep up	22.2	lay out	10.3
lead back	63.9	**get out**	**36.0**	turn back	22.1	set up	10.3
let down	63.3	**take off**	**36.0**	screw up	21.6	play out	9.9
make over	63.3	sort out	35.9	pass on	21.3	give up	9.7
set back	60.4	pull off	35.5	back up	21.1	open up	8.7
send back	57.4	hold back	35.4	sum up	21.0	go up	8.4
cheer up	57.1	win over	34.8	**check out**	**20.9**	take over	7.8
bring back	**56.9**	pay back	34.7	wrap up	20.9	break up	7.8
put down	**56.8**	think through	34.4	hold up	20.8	pay off	7.6
knock off	55.7	write down	33.8	lock up	20.5	build up	7.3
take in	55.0	drive back	33.8	slow down	20.5	turn up	6.7
drag down	54.8	turn around	33.7	cover up	20.5	go down	6.1
give back	54.2	make out	33.4	bring out	20.4	look down	5.9
draw in	54.0	push back	33.2	mess up	20.2	sit down	5.6
want out	53.6	finish off	33.0	tie up	19.9	take on	5.3
turn in	53.5	hold down	33.0	bring in	19.2	come down	4.9
help out	53.2	rip off	33.0	get up	19.0	point out	4.6
push down	53.1	mix up	32.9	get on	19.0	go out	4.5
turn on	51.2	try out	32.3	**pick up**	**18.9**	carry out	3.6
let out	48.2	put out	32.3	hold out	18.2	find out	3.2
put off	48.2	pass by	32.2	fill up	18.1	come out	2.3
drag out	47.9	pull back	32.0	call out	18.0	go back	2.0
put in	47.8	knock out	31.6	cut down	17.7	come up	1.4

Bolded = top 10 highest frequency SPVs.

4.2.1. Frequency of parts of speech in 4-g SPVs (research question #4)

Table 6 shows that only seven different grammatical options for slots 2 and 3 account for over 80% of all variation in the intervening slots of 4-g SPVs. The high frequency examples for each category suggest how ingrained these structures are in English, with many examples of both literal SPVs (*put this book down, put some clothes on*, etc.) and figurative SPVs (*put other people down, make it all up*, etc.).

Table 6 also suggests that one of the prime reasons for much higher variability in 4-g SPVs than in 3-g SPVs is that 4-g draw heavily from the large open class of nouns to fill in the slot immediately preceding the adverbial particle (slot 3), rather than a closed set of pronouns, as is the case with the 3-g forms. Expanding to slot 2, we see that most 2 + 3 slot constructions in 4-g SPVs are simple noun phrases with a common noun in slot 3 and either articles (38.4%), possessive determiners (28.2%), demonstrative determiners (6.0%), adjectives (2.0%) or quantifying determiners (1.6%) accounting for over 75% of slot 2 variability. We purposely did not collapse these noun phrases into one category because additional frequency information and examples can be gleaned from the more nuanced description, but we have grouped them together in the table to emphasize

Table 5
Top 50 highest frequency 3-g, 4-g, and 5-g SPVs.

RNK	3-G SPV	#	CUM %	4-G SPV	#	CUM %	5-G SPV	#	CUM%
1	pick up	22,006	2.5	get out	10,910	2.9	get out	2051	1.9
2	bring back	13,832	4.0	take off	8155	5.0	take out	1670	3.4
3	take out	11,429	5.3	take back	7019	6.8	get back	1498	4.8
4	**check out***	10,092	6.5	put down	6981	8.7	get up	1329	6.0
5	take back	9809	7.6	get back	6905	10.5	bring back	1317	7.2
6	**figure out***	9471	8.6	take out	5852	12.0	take off	1227	8.4
7	get back	9466	9.7	bring back	5346	13.4	take back	1161	9.4
8	set up	9291	10.8	put back	4782	14.7	put on	949	10.3
9	help out	7661	11.6	put up	4729	15.9	put back	896	11.1
10	get out	7571	12.5	get up	4186	17.0	put down	885	12.0
11	bring up	7232	13.3	turn around	3626	18.0	get in	874	12.8
12	put down	7111	14.1	put on	3357	18.8	put in	847	13.5
13	put back	7023	14.9	put out	3251	19.7	get down	845	14.3
14	make up	6677	15.6	keep out	3224	20.5	give up	843	15.1
15	**work out***	5874	16.3	make up	3063	21.3	make up	839	15.9
16	**look up***	5795	17.0	keep down	3001	22.1	put up	801	16.6
17	give up	5675	17.6	get on	2951	22.9	bring down	774	17.3
18	sort out	5127	18.2	bring down	2949	23.7	**go back***	713	18.0
19	take off	5099	18.7	get down	2756	24.4	get on	595	18.5
20	keep up	5044	19.3	put in	2667	25.1	take down	593	19.1
21	let down	4989	19.9	get in	2651	25.8	put out	578	19.6
22	turn around	4772	20.4	keep up	2576	26.5	**see out***	561	20.1
23	put out	4714	20.9	pick up	2473	27.1	**look down***	540	20.6
24	**put off***	4669	21.5	turn over	2435	27.8	**repeat over***	536	21.1
25	**point out***	4550	22.0	pull out	2372	28.4	get off	518	21.6
26	let in	4479	22.5	turn off	2285	29.0	give back	513	22.1
27	**cut off***	4446	23.0	take down	2129	29.5	keep down	508	22.5
28	**wake up***	4430	23.5	**hold up***	2103	30.1	turn off	507	23.0
29	put on	4358	24.0	get off	2035	30.6	keep up	470	23.4
30	give back	4344	24.5	let down	2001	31.2	turn over	458	23.9
31	**hold back***	4259	25.0	bring up	1955	31.7	**say over***	457	24.3
32	send back	4189	25.4	make back	1834	32.2	let out	448	24.7
33	**back up***	4159	25.9	throw out	1828	32.6	turn around	447	25.1
34	**pull off***	4149	26.4	**keep on***	1700	33.1	pull out	430	25.5
35	take in	4111	26.8	**stick out***	1610	33.5	throw out	391	25.9
36	bring down	4055	27.3	let out	1603	33.9	bring up	390	26.2
37	take down	3983	27.7	help out	1490	34.3	**see up***	382	26.6
38	**write down***	3962	28.2	send out	1483	34.7	make back	374	26.9
39	get down	3954	28.6	send back	1453	35.1	send back	369	27.3
40	get up	3871	29.1	take in	1452	35.5	**give out***	363	27.6
41	turn off	3848	29.5	set up	1442	35.8	keep out	351	27.9
42	put up	3729	29.9	**turn back***	1366	36.2	send out	348	28.2
43	take up	3698	30.3	make out	1364	36.6	**go down**	339	28.5
44	throw out	3582	30.7	bring in	1341	36.9	**see over***	339	28.9
45	**turn down***	3383	31.1	**pull up***	1316	37.3	make out	337	29.2
46	**go out***	3363	31.5	**work off***	1309	37.6	bring in	333	29.5
47	**slow down***	3339	31.9	**work up***	1301	37.9	take up	315	29.8
48	make out	3316	32.3	sort out	1289	38.3	let in	302	30.0
49	**try out***	3286	32.6	**set down***	1288	38.6	**give off***	297	30.3
50	put in	3187	33.0!	**shut up***	1286	39.0!	**pull down***	294	30.6!

*Unique in top 50 SPV 3-g, 4-g, and 5-g lemmas.

that they have the same basic noun phrase structure. The other major grammatical structures in 4-g SPVs are personal pronouns (slot 2)+ adverbs (slot 3), which account for 3.3% of 4-g SPV occurrences, and the reciprocal pronoun *each other* (slots 2 + 3) accounting for nearly 1%.

Perhaps the consistency and overall coverage of the grammatical structures in Table 6 could be a starting point for addressing 4-g SPV instruction in English language education, using higher frequency examples such as those in Table 6 for examination and practice. Certainly, raising awareness of which structures tend to intervene between a verb and its allied particle could give learners one more tool for dealing with SPVs in English.

4.2.2. Frequency of parts of speech in 5-g SPVs (research question #4 cont.)

Table 7 indicates that eight grammatical options for slots 2, 3, and 4 account for over 50% of all 5-g SPVs. While variability is higher for the top 5-g intervening grammatical structures (i.e., eight categories cover 51.9%) than for the 4-g (i.e., seven categories cover 80.3%), we would argue that knowing over half of the most likely grammatical structures may still be

Table 6
Most frequent parts of speech in slots 2 and 3 of 4-g SPVs.

Cat	Slot 2	Slot 3	% Tot	Cum%	Examples		
1 NP	Article (*the, a, an, no, every*)	Noun	38.4	38.4	take a step back	put the phone down	take a year off
					keep an eye out	take a day off	take a look back
					get the word out	get the message out	take a look around
					keep the weight off	turn the ball over	take the day off
2 NP	Possessive Determiner (*my, your, our*,etc.)	Noun	28.2	66.6	work his way up	work your way up	get their hand(s) on
					make his way back	get your hopes up	find their way back
					take my hat off	make their way back	turn his life around
					get my hands on	keep his head down	keep your head up
3 NP	Demonstrative Determiner (*this, that, these, those, another*)	Noun	6.0	72.6	make this stuff up	figure this one out	pick this one up
					sit this one out	pick this book up	get that message out
					put this book down	turn this thing around	get another caller in
					get this country back	turn this country around	check this one out
4 NP	Adjective	Noun	2.0	74.6	take great pride in	try new things out	let other people down
					put other people down	bring new players in	bring new people in
					help other people out	bring empty home back	let other people in
					come full circle back	hear good things about	hear great things about
5 NP	Quantifying Determiner (*any, some*)	Noun	1.6	76.2	take some time off	put some people off	take any time off
					take some time out	need some time off	get some clothes on
					get some caller in	put some clothes on	put some effort in
					see any way out	clear some things up	get some time off
6	Personal Pronoun (*it, them, me*, etc.)	Adverb (*all*)	3.3	79.5	take it all in	make it all up	give it all up
					top it all off	write it all down	add it all up
					figure it all out	bring it all back	lay it all out
					sort it all out	sum it all up	get it all out
7	Reciprocal Pronoun (*each*)	Reciprocal Pronoun (*other*)	0.8	80.3	cancel each other out	hold each other up	tear each other down
					help each other out	size each other up	stare each other down
					beat each other up	kill each other off	pick each other up
					balance each other out	feel each other out	take each other out

valuable for pedagogy and computing. As with the 4-g SPVs, common nouns dominate the slot immediately preceding the adverbial particle, and most of the grammatical structures of the 2 + 3+4 intervening slots are various versions of simple noun phrases, with the noun preceded either by an article + adjective (18.3%–e.g., *take the easy way out*), an article + the first part of a compound noun (8.8%–e.g., *get the unemployment rate down*), a possessive determiner + an adjective (6.3%–e.g., *get his old job back*), the quantifying determiner *all* + an article (4.9%–e.g., *go all the way back*), a possessive determiner + the first part of a compound noun (4.4%, e.g., *get your heart rate up*), or an article + the quantifying determiners *few, several*, or *many* (3.7%–e.g., *take a few days off*).

The last two categories in the top eight for 5-g SPVs do not have a noun in slot 4, but they are very instructive, providing additional evidence why complex phrases need more attention in both language education and linguistic computing. Category 7 contains a personal pronoun (similar to 3-g SPVs) + a complex adverbial particle containing the conjunction *and* (e.g., *look me up and down, turn it on and off*). A query to find a verb followed by an adverbial particle (phrasal verb) would miss these structures altogether, and for two separate reasons: (1) because of the intervening personal pronouns in slot 2, which makes these SPVs; and (2) because the SPVs are complex in the adverbial particle structure. In fact, even a search for 3-g or 4-g SPVs would miss these complex 5-g structures and potentially lead to inaccurate counts and conclusions. For language learning, it seems useful to teach the complex particles like *up and down*, *over and over*, and *on and off* as lexical items in their own right, because they are so ubiquitous in the language.

Finally, Category 8 in Table 7 is an intriguing example of the flexibility inherent in many SPVs, with 4-g SPVs suddenly becoming 5-g SPVs by inserting simple adverbs to modify the adverbial particle (e.g., *turn the music off > turn the music completely off*). Again, this complexity has important ramifications for both the computational processing of text and for language learning, especially at the intersection of syntax and meaning. Are there differences in meaning, for example, between *turn off the music completely* (2-g SPV), *turn the music off completely* (4-g SPV), and *turn the music completely off* (5-g SPV). Also, it is interesting that most experienced speakers of English would find a fourth option (*turn off completely the music*) to be unacceptable. But why? Our point is that separability in phrasal verbs has subtle ramifications that are important in actual language use. By extension, linguistic computing should more accurately account for these structures than we typically see in the research. To further demonstrate the role of separability as being more than just an option for some phrasal verbs, we can study other examples from Category 8 in Table 7. For instance, *read a book straight through* seems like a common enough 5-g SPV, but most of us would agree that its possible variations of *read through a book straight* (2-g PV), *read through straight a book* (also 2-g PV), and *read a book through straight* (4-g SPV) are not acceptable. Only the 5-g SPV allows this use of the adverb *straight* to modify the expression—specifically, the adverbial particle. The same is true for the adverb *right* in *turn the heat right down* and other similar structures.

Whether such nuances are worth instructional effort is debatable, but we propose that any time spent analyzing SPVs with learners will raise critical awareness of these structures, especially if we work with higher frequency examples that have

Table 7
Most frequent parts of speech in slots 2, 3, and 4 of 5-g SPVs.

Cat	Slot 2	Slot 3	Slot 4	% Tot	Cum %	Examples	
1 NP	Article (*the, a, an, no every*)	Adjective	Noun	18.3	18.3	take the easy way out	let the right one in
						work the other way around	call the whole thing off
						seem a long way off	work the other way round
						make the whole thing up	take a little time off
2 NP	Article	Noun	Noun	8.8	27.1	leave the toilet seat up	take the lens cap off
						put the toilet seat down	get the cheque book out
						get the unemployment rate down	bring the glory days back
						push the us economy back	bend the cost curve down
3 NP	Possessive Determiner	Adjective	Noun	6.3	33.4	try your new tactic out	get their old job back
						stretch his long leg out	put your right foot in
						get his old job back	put their collective foot down
						sing their little heart out	stick their political neck out
4 NP	Quantifying Determiner (*all*)	Article	Noun	4.9	38.3	go all the way back	come all the way down
						go all the way down	go all the way up
						come all the way out	go all the way through
						date all the way back	stretch all the way back
5 NP	Possessive Determiner	Noun	Noun	4.4	42.7	get your heart rate up	pull your shoulder blade down
						put your seat belt on	keep your heart rate up
						keep your energy level up	pull your shoulder blade back
						get my heart rate up	get your thinking cap on
6 NP	Article	Quantifying Determiner (*few, several, many*)	Noun	3.7	46.4	take a few days off	take a few weeks off
						take a few steps back	take a few years off
						remember a few years back	live a few doors down
						take a few months off	get a few days off
7	Personal Pronoun (*him, them, me*, etc.)	Adverbial Particle (*up, over*, etc.)	Conjunction (*and*)	3.3	49.7	look me up and down	turn it on and off
						say it over and over	read it over and over
						look him up and down	repeat it over and over
						look her up and down	watch it over and over
8	Article	Noun	Adverb (*right, straight*, etc.)	2.2	51.9	pull the rug right out	push the weight straight up
						turn the heat right down	put the decision firmly back
						pop the bugger right out	read a book straight through
						drop the cable straight down	turn the music completely off

greater potential utility. Perhaps one key for raising learners' awareness is to help them recognize how SPVs are modified by native English speakers to create specific meanings, and to start noticing such structures as they engage with the language.

5. Conclusions and recommendations

Our purpose in this study has been to better understand the complicated class of English separable phrasal verbs, and to provide language education and linguistic computing with lists of the most prolific of these structures based on a very large corpus of American and British English. These structures are relatively ubiquitous in the language, but they pose problems at the lexical level because they are phrasal forms, often with multiple meanings—some literal and some figurative. Perhaps even more problematic is that they, unlike their standard phrasal verb counterparts, are also separated in syntax, likely requiring English learners to have much more sophisticated language skills to both recognize and produce them (cf. Kadia, 1988). This separability aspect of these structures likewise causes problems for corpus-based research and computer-based processing of language.

By providing lists in this study of the most frequent SPVs and their internal grammatical structures, we hope to advance the field, and welcome any attempts to repurpose this information for classroom instruction, materials development, or research. Additionally, several pedagogical implications follow from the findings of this study:

1. When choosing examples of SPVs for instruction, it would make sense for teachers to utilize those on the higher frequency SPV lists provided in this study because learners are much more likely to encounter them and produce them in acutal usage.
2. We recognize along with Siyanova-Chanturia and Martinez (2015) that reducing groups of complex phrases to lists of decontextualized items to be learned and stored holistically can be problematic, particularly given the variability that exists in SPVs. For this reason, we encourage users of these lists to find need- and task-appropriate contexts for language training purposes (e.g., asking for directions, making a sales pitch, reading a sports article). This needs-based approach will also help when deciding which meanings of multi-meaning SPVs to emphasize.

3. A critical part of raising learners' SPV awareness would be for teachers to point out SPVs in materials the learners are expected to negotiate, and to encourage learners to be more aware of SPVs during their own extensive reading and listening experiences, perhaps offering incentives to learners who bring examples to class.

Finally, we recognize several limitations of this study that could be addressed in future research. First, while some spoken English is contained in the corpora we used (e.g., COCA and BNC), the vast majority of language is text-based. Our study did not address possible frequency differences between written and spoken SPVs, nor do we consider variation that might exist across different registers of English (spoken, fiction, newspapers, magazines, academic), or across different varieties of English (British vs. American English; inner-circle vs. outer-circle English, etc.). Second, beyond a simple analysis of possible word senses in high frequency SPVs, we made no attempt to analyze the various literal and figurative meanings among the highest frequency cases of SPVs, or with SPVs in general. Third, like all corpus-based work, there is an assumed margin of error in the data that could affect some of the figures provided in this article, and our liberal definition of SPV (any separation of a lexical verb from its allied adverbial particle) is certainly open to scrutiny. Finally, we fully acknowledge that the data we gathered is based on the particular corpora we chose to analyze, and that other corpora may result in different findings. But we are confident that no other corpora currently available can provide better data on this construction than what we have presented here.

Despite these limitations, we believe this study to be the most comprehensive analysis of separable phrasal verbs to date, and we hope the data will provide a solid starting point for dealing with these important structures in language education and the research that supports it. We also challenge other researchers to perform independent studies of these structures to provide additional insights, or to modify our conclusions.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.system.2018.06.009.

References

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

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

Cambridge Phrasal Verbs Dictionary. (2006) (2nd ed.). Cambridge: Cambridge University Press.

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

Collins Cobuild Phrasal Verbs Dictionary (2012). (London: Collins Publishers).

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly and nonformulaic language by native and nonnative speakers? *Applied Linguistics, 29*(1), 72–89.

Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics, 32*, 45–61.

Cornell, A. (1985). Realistic goals in teaching and learning phrasal verbs. *International Review of Applied Linguistics in Language Teaching, 23*(4), 269–280.

Dagut, M., & Laufer, B. (1985). Avoidance of phrasal verbs: A case of contrastive analysis. *Studies in Second Language Acquisition, 7*(1), 73–79. https://doi.org/10.1017/S0272263100005167.

Darwin, C. M., & Gray, L. S. (1999). Going after the phrasal verb: An alternative approach to classification. *Tesol Quarterly, 33*(1), 65–83.

Davies, Mark (2011). The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing, 25*, 447–s465.

Davies, Mark (2013). *Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries (GloWbE)*. Available online at https://corpus.byu.edu/glowbe/.

Davies, Mark, & Fuchs, Robert (2015). Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE). *English World-Wide, 36*, 1–28.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20*(1), 29–62.

Garnier, M., & Schmitt, N. (2015). The PhaVE list: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research, 19*(6), 645–666.

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

Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

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

Israel, M. (1996). The way constructions grow. In A. Goldberg (Ed.), *Conceptual structure, discourse and language* (pp. 217–230). Stanford, CA: CSLI Publications.

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

Kadia, K. (1988). The effect of formal instruction on monitored and on spontaneous naturalistic interlanguage performance: A case study. *Tesol Quarterly, 22*(3), 509–515.

Kremmel, B., Brunfaut, T., & Alderson, J. C. (2017). Exploring the role of phraseological knowledge in foreign language reading. *Applied Linguistics, 38*(6), 848–870.

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*(1), 35–48. https://doi.org/10.1017/S0272263100011657.

Leech, G. (2007). New resources, or just better old ones? The Holy Grail of representativeness. In M. Hundt, N. Nesselhauf, & C. Biewer (Eds.), *Corpus linguistics and the web* (pp. 133–149). Amsterdam: Rodopi.

Leech, G., & Smith, N. (2000). *Manual to accompany The British National corpus (version 2) with improved word-class tagging*. Lancaster: Lancaster University.

Liao, Y., & Fukuya, Y. J. (2004). Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning, 54*(2), 193–226.

Liu, D. (2011). The most frequently used English phrasal verbs in American and British English: A multicorpus examination. *Tesol Quarterly, 45*(4), 661–688.

Longman Phrasal Verbs Dictionary (2nd ed.) (2000). (London: Pearson Longman).

Master, P. (1996). *Systems in English grammar*. Englewood Cliffs, NJ: Prentice Hall Regents.

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

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 University Press.
NTC's Dictionary of Phrasal Verbs and Other Idiomatic Phrases (1993). (Chicago: National Textbook Company).
Oxford Phrasal Verbs Dictionary for Learners of English (2nd ed.) (2006). (Oxford: Oxford University Press).
Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards, & R. W. Schmidt (Eds.), *Language and communication*. London: Longman.
Princeton University. (2010). *About WordNet*. Princeton University. Accessible at http://wordnet.princeton.edu.
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*(1), 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*(1), 160–164.
Siyanova-Chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics, 36*(5), 549–569.
Trebits, A. (2009). The most frequent phrasal verbs in English language EU documents—a corpus-based analysis and its implications. *System, 37*, 470–481.
Widdowson, H. G. (2000). On the limitations of linguistics applied. *Applied Linguistics, 21*(1), 3–25.
Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics, 21*(4), 463–489.
Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.