System 59 (2016) 29–44

Contents lists available at ScienceDirect

System

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

# *Picking up* polysemous phrasal verbs: How many do learners know and what facilitates this knowledge?

Mélodie Garnier*, Norbert Schmitt

*School of English, University of Nottingham, Nottingham, NG7 2RD, UK*

ARTICLE INFO

*Article history:*
Received 19 May 2015
Received in revised form 7 April 2016
Accepted 8 April 2016
Available online 20 April 2016

*Keywords:*
Phrasal verbs
L2 acquisition
Polysemy
Frequency
Semantic transparency
L2 exposure

ABSTRACT

This study investigates L2 learners' knowledge of highly frequent polysemous phrasal verbs in English, and the effect of a number of factors on this knowledge. 128 students on BA English/TEFL courses were recruited to take a productive test in the form of a gap-fill task. The results show that only 40% of phrasal verb meaning senses were known on average, with the chances of knowing all the different meaning senses of each phrasal verb tested being quite low at only around 20%. The factors of semantic opacity, previous L2 instruction, immersion in L2 environment, and year of BA study did not have any effect on knowledge. Conversely, corpus frequency was found to predict knowledge, along with time spent reading per week, and time spent social networking per week. No relationship was found between phrasal verb knowledge and the hours spent listening to music and watching films in English. The study confirms that phrasal verbs are a problematic feature of English vocabulary for many learners, and consequently deserve more attention – either via instructed contexts or outside the classroom in L2 language engagement.



## 1. Introduction

It is now well-established that formulaic language is an essential part of the English lexicon, and thus should be known by L2 learners in order to produce fluent, competent language (Wray, 2002, 2008; Annual Review of Applied Linguistics 32, 2012). Phrasal verbs (hereafter PVs) are one category of formulaic language that is very common in English, especially in spoken discourse. In vocabulary studies, word frequency is roughly synonymous with usefulness; the more frequent a word in an L2, the more useful it is to know (Ellis, 2002; Nation, 2001). Therefore, there is little doubt that knowing highly frequent PVs is necessary for proficient language use, and that research is needed to gauge L2 knowledge of these items. However, PVs are highly polysemous (Gardner & Davies, 2007), and little research to date has taken this polysemy into account when studying L2 learner knowledge of PVs.

Furthermore, vocabulary research has begun to identify a number of factors which affect the learning of both individual words and formulaic sequences, most notably word frequency, semantic opacity, amount of L2 instruction, and degree of involvement/exposure with the language (Schmitt, 2010). Among the small number of studies which have looked at L2 knowledge of PVs, none has sought to assess the effect of these factors in combination. This study will address these gaps by

* Corresponding author.
*E-mail addresses:* melodie.garnier@nottingham.ac.uk (M. Garnier), norbert.schmitt@nottingham.ac.uk (N. Schmitt).

http://dx.doi.org/10.1016/j.system.2016.04.004
0346-251X/

investigating L2 learner knowledge of polysemous PVs, and by exploring the factors which relate to the learning of these various meaning senses.

## 2. Literature review

### 2.1. The importance of phrasal verbs in English vocabulary

There are several reasons why English PVs are important to learn. The first is that they have been found to be very frequent in everyday language. For example, based on a corpus search of the British National Corpus (BNC), Gardner and Davies (2007) estimate that learners will encounter, on average, one PV in every 150 words of English they are exposed to. Biber, Johansson, Leech, Conrad, and Finegan (1999) estimate that PVs occur almost 2000 times per million words. PVs are therefore an important component of English vocabulary. Because they are widely used in spoken informal discourse, failure to use them in such situations is likely to make language sound unnatural and non-idiomatic (Siyanova & Schmitt, 2007).

However, PVs are often considered to be one of the most challenging features of the English language. Firstly, they may be seen as an unnatural construction for some learners whose L1 lacks such a structure. Their syntactic peculiarity (some PVs allow for particle movement, others do not) and semantic complexity (some PVs have meanings that are highly idiomatic and opaque) make them particularly difficult to learn, and they are prone to avoidance (Dagut & Laufer, 1985; Hulstijn & Marchena, 1989; Laufer & Eliasson, 1993). In addition, they are composed of two or more orthographic words, which means that instead of recognising them as single semantic units, unaware learners may attempt to decode the meanings of their individual components, and therefore misinterpret them.

### 2.2. L2 knowledge and use of phrasal verbs

Some previous research has suggested that because PVs are so challenging, L2 speakers typically use one-word verbs that are more or less synonymous instead (e.g. *postpone* instead of *put off*). For example, Siyanova and Schmitt (2007) found that advanced learners showed a strong preference for one-word verbs whereas native speakers showed a preference for PVs in about half of the cases. A number of studies have observed an avoidance phenomenon in PV use, which can be defined as the conscious decision from learners not to use a particular L2 form, although that L2 form is known (Dagut & Laufer, 1985; Hulstijn & Marchena, 1989; Laufer & Eliasson, 1993; Liao & Fukuya, 2004).

In addition, a number of studies have sought to quantify and compare the use of PVs in learner corpora as opposed to native corpora. For instance, Waibel (2007) found that the frequency of PVs in many sub-corpora of the ICLE (e.g. French, Italian, Spanish, Russian) is lower than that of the LOCNESS (a native corpus). A recent corpus study (Chen, 2013) explored Chinese university students' use of PVs in comparison with their American and British counterparts by comparing a corpus of learner English and four native corpora of two English varieties and two genres (argumentative and academic writing). The author concluded that, in general, the learners did not show a numerical difference from the native writers in PV use. However, this does not mean that the Chinese students' knowledge of PVs reached a native level, for at least two reasons. Firstly, as acknowledged by the author, their frequent use of PVs may be due to their highly frequent use of verbs in general. Secondly, the study does not tell us the distribution of PV use. It may be that some PVs were overused by being repeated again and again, even across writers.

To our knowledge, only two studies have directly measured L2 learners' knowledge of PVs. For example, Schmitt and Redwood (2011) found that most of their participants were able to recognise most of their tested PVs (65.2%) receptively, and about half of them productively (48.2%), which implied that participants had relatively good knowledge of the selected PVs considering their intermediate level of English. However, studies like Schmitt and Redwood have typically only looked at a single PV meaning sense. Unfortunately, this does not fit well with the finding by Gardner and Davies (2007) that PVs are highly polysemous. They found that the 100 most frequent English PVs as identified by their BNC search had 5.6 meaning senses on average. Future studies into knowledge of PVs will need to take this polysemy into account, in order to get a better account of learners' true knowledge of PVs.

### 2.3. The role of frequency in the acquisition of PVs

Frequency has long been considered an essential predictor of L2 vocabulary knowledge, i.e. the more frequent a word, the more likely it is to be known (Schmitt, 2010). It is widely recognized that learners generally acquire higher-frequency words before lower-frequency ones (Ellis, 2002; Leech, Rayson, & Wilson, 2001; Nation, 2001; Nation & Waring, 1997). The question that thus arises is whether the same holds true for multi-word combinations such as PVs. However, because research interest in formulaic sequences is much more recent than research interest in vocabulary in general, the strength of the relationship has not been extensively documented. The few studies focusing on the relationship between frequency and knowledge of formulaic language typically involve collocations (e.g. Siyanova & Schmitt, 2008).

In reality, with the exception of tightly-controlled experiments where it is possible to either know or control for the number of exposures any learner receives, there is no way of knowing how many times any specific word or PV is encountered

by any learner in most learning contexts. Nevertheless, for convenience and practicality reasons, corpora have typically been used as the primary indicator of frequency. Because they rely on computers which allow for fast and accurate counting, corpus counts are objective and quantifiable, and thus the best tool for uncovering language patterns which could otherwise be difficult to intuit (Reppen & Simpson-Vlach, 2010).

Schmitt and Redwood (2011) compared the results of 68 EFL/ESL students on productive and receptive tests with PV frequency rankings from the BNC complete, BNC written, and BNC spoken, and subsequently the COCA; they found significant positive correlations between mean test scores and PV frequencies (receptive = 0.29–0.31; productive = 0.41–0.46). The strength of the correlation coefficient was deemed to be fairly strong for the productive test, but moderate for the receptive test. For the BNC complete, frequency accounted for about 20% of the variance in the productive scores, and for only 9% of the variance in receptive scores. For the COCA complete, frequency accounted for around 18% of the variance in productive scores, and for 13% of the variance in receptive scores. Thus, the results are rather similar across the two corpora. The authors therefore conclude that frequency (as indicated by corpus data) seems to predict the productive scores of their participants to a considerable degree (but not so much their receptive scores).

Following Schmitt and Redwood, Chen (2013) also sought to ascertain the link between PV knowledge and corpus frequency. She examined her Chinese EFL learners' production of PVs, in light of the frequency rankings of the 50 most frequent PVs in the BNC and the COCA. She found that her participants' production of PVs had a positive correlation with PV frequency, with the $r^2$ variance being greater with the COCA (17.1%) than with the BNC (11.9%).

Albeit informative, one crucial limitation of these two studies is the fact that the frequency of the PVs was taken as the number of occurrences of PV forms in a corpus, regardless of how many different meaning senses they entailed. As we have previously seen, polysemy is a key feature of PVs. Therefore, we might suspect that the strength of frequency as a predicting factor of PV knowledge would be even greater if polysemy was taken into account.

### 2.4. The role of semantic opacity in the acquisition of PVs

In the field of pedagogy, the Conceptual Approach (Dirven, 2001; Kurtyka, 2001; Rudzka-Ostyn, 2003) is based on the assumption that PVs have prototypical (literal) meanings from which metaphorical (figurative) meanings have been derived over time; this explains their polysemic nature. For this reason, it makes little sense to talk about 'literal PVs' or 'figurative PVs'. Instead, we should talk about 'literal meaning senses' or 'figurative meaning senses'. The question that arises is whether literal meanings are easier to learn, and thus better known, than figurative ones. An intuitive assumption is that, for any formulaic item, literal meanings should be better known than figurative ones since they can be easily inferred from their individual components. The PHRASE List compiled by Martinez and Schmitt (2012), which includes many PVs, was fundamentally based on this assumption. Among the core criteria used by the authors to determine inclusion or non-inclusion of phrasal expressions in their list was semantic opacity: all items present in the list were identified as semantically opaque, thus potentially causing difficulty for learners of English.

However, to date, no study has sought to directly investigate the effect of semantic opacity on PV knowledge. The issue of semantic opacity has mostly been discussed in relation to idioms, as idioms are often perceived as formulaic items at the extreme end of the 'opaque' spectrum (e.g. *bite the bullet*) (see Conklin & Schmitt, 2012, for a review of research on the processing of formulaic language – mainly idioms). Furthermore, other factors might complicate the issue. For example, Conklin and Schmitt (2008) found that figurative interpretations may be more frequent than their literal counterparts in some formulaic sequences, leading to a tension between the facilitating effects of frequency and the inhibiting effects of idiomaticity. Since frequency tends to be a major factor in predicting acquisition, it is still unclear whether the learning of PVs is more affected by frequency or semantic opacity.

### 2.5. The role of everyday engagement with the L2 in the acquisition of PVs

We have reviewed two characteristics of PVs (corpus frequency and semantic opacity) as they relate to learning. But if corpus frequency may only relate moderately to learning (Schmitt & Redwood, 2011), might a learner's personal exposure relate more strongly? This concerns the learner's degree of communicative engagement with a second language. Communicative engagement includes activities where the L2 is used with a specific purpose of a social or leisurely nature, e.g. reading books, watching films, listening to music, and using social media. Spending time in an L2 environment is typically conducive to a very high degree of communicative engagement.

Schmitt and Redwood (2011) found that both extensive reading and watching English language films and television had a significant positive relationship with PV knowledge, but that listening to music in English or using English social networking sites extensively did not. González Fernández and Schmitt (2015) tested Spanish speakers' productive knowledge of English collocations in relation to various factors, including degree of engagement with the L2 via reading, watching TV, listening to music, and using social media. They also found that extensive reading and watching films/TV had significant positive relationships with collocation knowledge, but not listening to music. However, they additionally found a significant positive relationship between using English social networking sites and knowledge of their target collocations. These two studies

suggest that using English for leisurely activities outside the classroom significantly contributes to better knowledge of formulaic sequences (PVs and collocations).

Conversely, the effect of spending time in the L2 environment on L2 knowledge seems to be more moderate. Siyanova and Schmitt (2007) found that the length of time spent in L2 environments did not have a strong effect on their participants' likelihood of using PVs instead of one-word verbs. A more relevant predictor of formulaic knowledge, however, seems to be the quality of exposure in the L2 environment (i.e. how often they interact with native speakers and the L2 culture as a whole), or sociocultural adaptation. Both Adolphs and Durow (2004) and Dörnyei, Durow, and Zahran (2004) observed a significant positive relationship between the degree of sociocultural adaptation and mastery of formulaic sequences.

The above review of the literature showed that previous research into PV knowledge did not take account of polysemy, and that it is unclear how semantic frequency (as indicated by corpus data) and semantic opacity affect PV learning. There is some evidence that personal leisure exposure relates to formulaic sequence learning, but this has been shown only with a small group (68) of intermediate/upper intermediate English learners for PV learning (Schmitt & Redwood, 2011) and for the learning of collocations (González Fernández & Schmitt, 2015). In order to better understand the knowledge and learning of PVs, this study asks the following questions:

1. How good is L2 learners' knowledge of highly frequent PVs and their various meaning senses?
2. What factors can be identified as the best predictors of PV knowledge (among linguistic factors such as frequency and semantic opacity, and exposure factors such as L2 instruction, L2 immersion and L2 engagement in leisure activities)?

## 3. Methodology

### 3.1. Participants

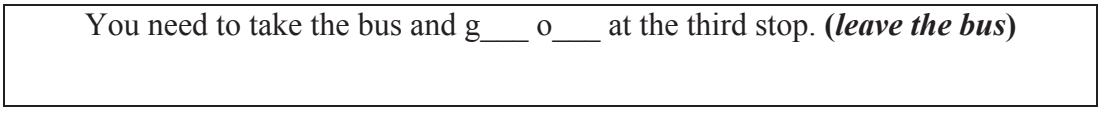
The participants were 128 Chilean students of English (36 males, 84 females, 8 unknown) from two Chilean universities. The age range was 18–44 ($M = 22.6$; $SD = 3.7$). At the time of data collection, they were all following a mixed English-medium and Spanish-medium BA course in either English Language and Literature or TEFL in their respective universities. In order to cope with the English-medium component, we initially presumed that they had a relatively high level of proficiency, but this assumption could not be confirmed as we did not have information on their scores in a recognized proficiency test. In an attempt to account for the effect of year of study on their PV knowledge, these students were recruited from 1st to 4th year of study in each university in roughly equal numbers (1st-year = 27, 2nd = 31, 3rd = 40, and 4th = 30). Finally, all participants shared the same L1, i.e. Spanish. Consequently, any score variation could not be attributable to differences in the L1.

### 3.2. Selection of target items

PVs are highly polysemous (Gardner & Davies, 2007), and we wanted our study to reflect this. We thus used items from the PHaVE List, which presents the most frequent meaning senses of the most frequent English PVs (Garnier & Schmitt, 2015). The list includes the 150 PVs which have been identified by previous research (Liu, 2011) as being the most frequent, i.e. having at least 10 tokens per million words in either the COCA or the BNC. The most frequent meaning senses were extracted from the COCA, on the basis of accounting for at least 10% of the PV occurrences in the whole corpus. A total of 288 meaning senses are listed (two per PV on average), so not all of these could be included on our test due to time restrictions. We opted for a 33% sampling rate (i.e. 50 PVs) for our first version of the test at the onset of the piloting stage. In order to avoid bias in the process of selecting items, they were drawn from the whole PHaVE List via a random number generator; they could thus be assumed to represent the complete range of frequencies represented by the PHaVE List, along with varying degrees of semantic opacity. The frequencies of each PV meaning sense were drawn from the PHaVE List, which provides this information. Semantic opacity (L = literal: *I **took off** my shirt and went to bed*; F = figurative: *They jumped into the car and **took off***) was established by the authors' judgments, in conjunction with another native speaker of English.

### 3.3. Test format

The instrument for data collection was an off-line pencil-and-paper *form recall* test in the form of gap-fill sentences. Each item was embedded in a sentence in English (the participants' L2), setting the context and serving as a prompt. Each sentence comprised two gaps, corresponding to each of the two words which form the PV meaning sense tested (lexical verb and particle). To help the participants and constrain the range of potential PVs elicited, the first letter of each of the two words was given. At the end of each sentence was the meaning of the elicited PV in brackets, printed in bold and italics in order to make it more noticeable. Below is an example of a sentence included in the test:



Since the PVs on the test were polysemous, we ensured that various meaning senses for a PV were effectively spread out across the test in order to avoid priming and facilitation effects. Participants were informed that some sentences used the same PV as others (so as to avoid potential confusion), and that they were allowed 30 min to complete the test. All instructions, along with information sheets and consent forms, were written in the participants' L1 (i.e. Spanish) in order to maximise comprehension and efficiency. The Spanish translations were reviewed and approved by a native speaker of Spanish.

### 3.4. Test piloting

A series of pilot tests, comprising two stages, was conducted in order to check the validity of the test. The first stage was the administration of the pilot tests to 30 native-speaking university students, to check whether the items on the test could be correctly answered from the sentence contexts and given meanings, and to get an idea of the time needed to complete the test. The results showed that correct answers were provided for the large majority of items, and test-taking time was usually around 25 min. Items for ten PVs proved problematic, and were discarded from the revised version, which was then piloted in the second stage with three non-native-speaking PhD students with three different L1s (French, German, and Arabic).

As with the native speakers, the students were allowed 50 min and the possibility of asking questions if they wished. This time however, a greater score variation was expected due to the fact that non-native speakers are more likely to differ in their vocabulary knowledge. Their scores were 48% (L1: French), 62% (L1: Arabic), and 97% (L1: German). The distribution of scores gave us confidence that the test was tapping into varying degrees of PV knowledge. After taking the test, the participants indicated that they felt comfortable with the procedure, and that they did not notice any confusing items. Test-taking time ranged from 20 to 35 min including review and signing of consent forms. Following this, the test was deemed adequate and ready to be administered to other non-native speakers of English, with a completion time allowance of 40 min in total (30 min for completing the PV test and 10 for completing the consent form and reading the information sheet).

The final test included 40 PVs (comprising 100 meaning senses). They can be found on Appendix A in alphabetical order, with meaning sense frequencies in decreasing order and semantic opacity category. The definitions are taken from the PHaVE List.

### 3.5. Questionnaire

In order to determine the effect of a number of factors on PV knowledge, a biodata questionnaire was included at the end of the test. The principle behind this questionnaire is that participants have different characteristics which inevitably affect lexical knowledge (Schmitt, 2010). It is thus useful to gather biodata information in order to account for score variation.

One factor that was of particular interest was everyday exposure to English via reading, listening to music, watching films, and using social media. In today's world where social media and communication permeate our lives, these activities are undeniably gaining popularity, especially among younger learners and students. In addition, such activities appear to be particularly suitable for acquiring PVs, since PVs are most commonly found in informal discourse. The importance of everyday engagement in vocabulary acquisition was clearly evidenced in our literature review. Therefore, participants were asked to estimate the number of hours they spent every week on each leisure activity. Several questions relating to previous L2 exposure in and outside classrooms (years spent studying general English and months spent in countries where the L2 is spoken) were also included. The purpose of the questionnaire, along with brief instructions, was presented before the questions (in the participants' L1 to maximise comprehension).

Among the 128 participants, 18 left their questionnaires blank or provided some unrealistic responses, which means that their data was discarded in the second part of the analysis (RQ 2).

### 3.6. Test administration

The tests were administered in Chile, on the premises of the two universities, and under the supervision of at least one member of teaching staff. They were completed on paper and pencil form, with the 30 + 10 time format mentioned above. The reason for setting a time limit was to prevent guessing behaviours from participants, because all PVs were used for multiple meaning senses. Prior to taking the test, all participants were given explanations as to the aim of the study, the format of the test, its completion time, voluntary nature, and the confidentiality of the data. An English version of our test and questionnaire can be found on Appendix B.

### 3.7. Data analysis

The maximum score for each test was 100, based on 1 point per correct meaning sense (both words – lexical verb and particle – had to be correct). Accurate spelling was not necessary for the item to be judged correct, as long as the intended answer was clear. Likewise, following Schmitt and Meara (1997), inflectional mistakes were not considered. Once test scoring was completed, the data was imported into Excel spreadsheets and IBM SPSS for analysis. In order to answer our second

research question, we decided to use mixed-effects models because they allow for the inclusion of both subject and item as random effects in addition to fixed effects. This allows the researcher to account for individual differences in subjects as well as in items. It also eliminates the need for separate analyses with participants and items as random variables. The results were analysed using an omnibus linear mixed effects model with the lme4 package (version 1.1–10, Bates et al., 2014) in R (version 3.2.2, R Core Team, 2014).

## 4. Results and discussion

### 4.1. How good is L2 learners' knowledge of highly frequent PVs and their various meaning senses?

Schmitt and Redwood (2011) reported that their 68 intermediate/upper intermediate participants had reasonably good knowledge of PVs ($M = 48\%$) on a gap-fill test similar to ours. However, they tested only one meaning sense per PV, typically the most frequent one. If we consider only the most frequent meaning sense of each PV, our results were lower, with an average of 44.5% being known. In absolute terms, it seems that knowing less than 45% of the most frequent meaning senses of the most common PVs is not a very strong result for students successfully working at a BA level. It suggests that learning English mainly in an instructed setting, even if successfully (i.e. learners made it to the BA level), does not guarantee acquisition of elements of language which are mainly informal in nature and might be acquired better from less formal contexts. We will have more to say about this later.

Beyond the most frequent meaning sense, our study is able to provide a much more detailed description of learners' PV knowledge than previous studies. Looking at all of the 100 meaning senses tested, participants were able to produce the correct meaning sense for 40.56 on average (Table 1). Again, this is relatively modest given that our participants were BA students, and far less than half of the items. The spread of correct answers ranges from 4 to 93 (i.e. 4%–93% on the 100-item test). The vast majority of participants (N = 94) scored between 20 and 60%. Around 12% (N = 15) participants scored less than 20%, whilst around 15% (N = 19) scored more than 60%.The distribution of test scores is illustrated in Fig. 1, with scores grouped in 5s. The figure displays a bell curve, but results are visibly skewed towards the lower half.

Finer grained information can be obtained by looking at our participants' knowledge of the individual meaning senses of each PV form. Table 2 illustrates knowledge of up to the 4th most frequent meaning sense of each target PV. The means for each sense are: 1st most frequent – 44.5%, 2nd – 40.1%, 3rd – 31.6%, and 4th – 44.3%. Overall, it seems that the knowledge does not drop in parallel to coverage ranking as the means of the 4th meaning sense is nearly as high as that of the 1st meaning sense (however, it is worth noting that only 3 PVs in the test had four meaning senses, and therefore our means for the 4th meaning sense may not be as reliable as the others). On closer inspection of the table, we can see that for most PVs, the percentage known bounces up and down among the coverage ranks with no clear pattern. In some cases, knowledge drops with coverage rank in an expected way (e.g. *take in*: 64.8%↘28.9%↘22.7%; *put in*: 78.9%↘43%), but there are nearly as many cases where the opposite trend occurs (e.g. *make out*: 10.9%↗17.2%↗28.9%; *sit back*: 52.3%↗71.9%). Also, for the three PVs whose PHaVE List coverage percentages were very close (within 5%), the knowledge percentages were not close, but showed an inconsistent pattern (*cut off*: 46.1%↘25.8%↗61%; *put up*: 69.5%↘56.2%↗63.3%; *pay off*: 17.2%↗22.7%).

Overall, the means for each meaning sense are quite low, especially the 3rd most frequent sense. This builds an even stronger case for the difficulty of PVs, especially if one thinks of a comprehensive knowledge of multiple PV meaning senses. Out of 384 possible cases (128 participants × 3 PVs), in only 72 (18.75%) were all four meaning senses of *break down, come out,* or *hold back* known. For the PVs with three meaning senses, in only 11.27% of cases were all three known. Even for PVs with only two senses, both senses were known in only 29.18% of the cases. Overall, these results indicate that it makes little sense to speak of PVs being 'known' or not based on a single meaning sense. If one wishes to know whether learners know the multiple meaning senses, these need to be measured separately rather than just inferred from the score of a single (usually most frequent) meaning sense.

### 4.2. What factors can be identified as the best predictors of PV knowledge?

As previously mentioned, mixed-effects modelling was chosen to gauge the effect of our variables on PV knowledge. Because we had multiple predictor variables (including several usage variables which might be correlated), we first checked for any significant correlations. Since none of the usage variables were significantly correlated, we proceeded with the mixed effects analysis including all variables as potentially important fixed effects.

Our model development procedure was conducted in the following way. First of all, we log transformed COCA frequency in order to reduce skewing, as the data had very wide ranges (from 347 to 19,765 occurrences). Because our independent

**Table 1**
Descriptive statistics of the participants' test scores.

| | Minimum (%) | Maximum (%) | Mean (%) | Standard deviation |
|---|---|---|---|---|
| Test scores (N = 128) | 4 | 93 | 40.56 | 18.5 |

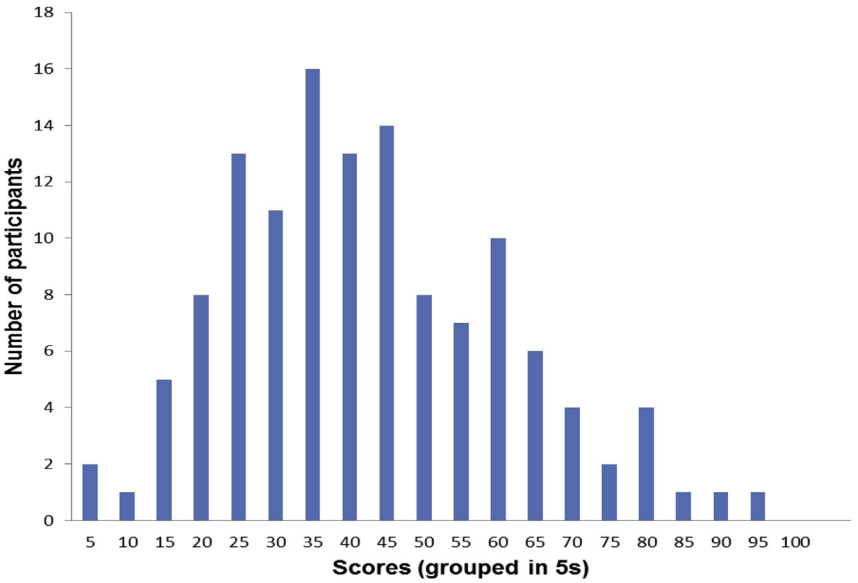



**Fig. 1.** Distribution of participants' test scores (grouped in 5s).

variable (knowledge) was binary, we used a generalised linear model with binomial regression. An initial model was built including all our explanatory variables as predictors of knowledge: (log) COCA frequency, semantic opacity, year of BA study, previous L2 instruction, previous immersion in L2 country, hours spent reading per week, hours spent watching films per week, hours spent listening to music per week and hours spent social networking per week, with participants and items as random variables. As expected, not all of the nine explanatory variables proved to be statistically significant.

We then proceeded by using a backwards stepwise procedure to eliminate variables that did not significantly contribute to the fit of the model. The process involved eliminating the variable with the lowest *z*-value and then refitting the model. This procedure continued until all insignificant variables were removed. The order of elimination of insignificant variables was the following: previous immersion in L2 country, hours spent watching films per week, previous L2 instruction, hours spent listening to music per week, year of BA study, and semantic opacity. Every time the model was refitted, it was compared to the previous one to confirm that including these variables did not significantly improve the overall model. As a result, the final best-fit model included three variables as significant predictors of knowledge: (log) COCA frequency, hours spent reading, and hours spent social networking per week. A summary of our results can be seen in the following table:

Finally, we wanted to check whether there was any evidence of an interaction between log COCA Frequency and semantic opacity. We therefore constructed a final model which included the fixed effects listed in Table 3, and included an interaction of Frequency and Opacity. There was no evidence of such an interaction ($z = -0.09$, $p = 0.93$), and explicit model comparison showed that inclusion of this interaction in the model did not significantly improve the fit ($\chi^2(2) = 3.12$, $p = 0.21$). To sum up, the omnibus analysis shows clear effects of COCA frequency, reading, and using social media on PV knowledge. Conversely, no effect was found for semantic opacity, previous L2 instruction, immersion in L2 country, year of BA study, watching films or listening to music.

*4.2.1. Linguistic factors: corpus frequency and semantic opacity*

Our test included 32 literal and 68 figurative meaning senses, with frequencies ranging from 347 to 19,765 occurrences ($M = 2991$; $SD = 2801$). According to our results, corpus frequency was clearly a strong predictor of PV knowledge ($p < 0.001$). This suggests that, similar to individual words, the importance of frequency as a predictor of L2 vocabulary knowledge also extends to PVs: the more frequent a PV, the more likely it is to be known. Our finding is congruent with Schmitt and Redwood (2011) and Chen (2013) who also deduced a positive relationship between frequency and productive knowledge of PVs in their studies. Compared to those, the present study has the additional advantage of involving semantic frequency counts (as opposed to rank frequencies of word form), which means that our frequency figures are likely to be more reliable and precise. In addition, using mixed-effects modelling allows us to identify predictors with confidence (whereas correlations are more difficult to interpret). In sum, whilst the effect of frequency on knowledge had been uncovered by previous studies, the fact that it was clearly demonstrated in our study confirms its fundamental importance in predicting PV knowledge.

Conversely, semantic opacity had no effect on PV knowledge in our study and was, in all likelihood, overridden by frequency. Also, it might be that semantically transparent items are less noticeable to learners, and therefore fail to be acquired.

**Table 2**
Knowledge of the various meaning senses of each PV form (40 PVs = 100 senses).

| PV | Coverage frequencies of meaning senses from the PHaVE list (%) [a] | | | | 1st most frequent meaning sense scores (%) | 2nd most frequent meaning sense scores (%) | 3rd most frequent meaning sense scores (%) | 4th most frequent meaning sense scores (%) |
|---|---|---|---|---|---|---|---|---|
| Break down | 24 | 20 | 17.5 | 13.5 | 68 | 39 | 47.7 | 26.6 |
| Come out | 38 | 13.5 | 11.5 | 10 | 61 | 37.5 | 42.2 | 73.4 |
| Hold back | 23.5 | 21 | 17.5 | 16 | 35.2 | 39 | 51.6 | 32.8 |
| Back up | 26 | 21 | 20.5 | | 31.2 | 48.4 | 36.7 | |
| Break off | 40 | 28 | 24 | | 11.7 | 8.6 | 13.3 | |
| Cut off | 27 | 24.5 | 23.5 | | 46.1 | 25.8 | 61 | |
| Get down | 26 | 22.5 | 17.5 | | 18 | 51.6 | 28.1 | |
| Get off | 54 | 12.5 | 12 | | 30.5 | 3.1 | 4.7 | |
| Give out | 40 | 33.5 | 11.5 | | 25 | 36 | 7.8 | |
| Go down | 29 | 27 | 18 | | 73.4 | 82.8 | 32.8 | |
| Make out | 60.5 | 11 | 10.5 | | 10.9 | 17.2 | 28.9 | |
| Put out | 47 | 14 | 10 | | 25 | 25.8 | 20.3 | |
| Put up | 23 | 19 | 18 | | 69.5 | 56.2 | 63.3 | |
| Set out | 42.5 | 26.5 | 16 | | 17.2 | 18.7 | 25 | |
| Take in | 24.5 | 17.5 | 10 | | 64.8 | 28.9 | 22.7 | |
| Take out | 50.5 | 13.5 | 12.5 | | 41.4 | 50.8 | 13.3 | |
| Turn up | 48 | 21.5 | 14 | | 17.2 | 93 | 37.5 | |
| Bring in | 52 | 30.5 | | | 65.6 | 37.5 | | |
| Clean up | 74 | 22 | | | 71.1 | 20.3 | | |
| Come along | 72.5 | 20.5 | | | 24.2 | 32 | | |
| Come in | 65 | 14 | | | 89.8 | 75 | | |
| Come on | 50 | 19.5 | | | 93 | 79.7 | | |
| Get on | 51 | 14.5 | | | 5.5 | 45.3 | | |
| Go up | 47.5 | 20.5 | | | 57 | 56.2 | | |
| Hand over | 58.5 | 41.5 | | | 7.8 | 6.2 | | |
| Keep up | 46 | 32.5 | | | 49.2 | 79.7 | | |
| Look back | 49.5 | 30 | | | 57.8 | 64.8 | | |
| Look out | 50.5 | 25.5 | | | 71.9 | 53.9 | | |
| Move up | 47 | 22.5 | | | 53.9 | 39.8 | | |
| Pay off | 49 | 48.5 | | | 17.2 | 22.7 | | |
| Pull back | 66.5 | 31 | | | 21.1 | 20.3 | | |
| Put in | 50 | 26.5 | | | 78.9 | 43 | | |
| Put on | 52 | 14.5 | | | 84.4 | 14.1 | | |
| Reach out | 48.5 | 39.5 | | | 28.9 | 24.2 | | |
| Run out | 49.5 | 34 | | | 57.8 | 51.6 | | |
| Sit back | 66 | 34 | | | 52.3 | 71.9 | | |
| Stand out | 60.5 | 38 | | | 30.5 | 10.9 | | |
| Take back | 50 | 33.5 | | | 53.1 | 53.1 | | |
| Turn around | 67.5 | 24.5 | | | 52.3 | 11.7 | | |
| Turn over | 59.5 | 34 | | | 11.7 | 28.9 | | |
| **AVERAGE** | **48** | **25** | **15.5** | **13** | **44.5** | **40.1** | **31.6** | **44.3** |

[a] Because the PHaVE List focused on the most frequent meaning senses which made up at least 75% of the total PV occurrences, not all possible meaning senses were measured. Thus the totals will add up to less than 100%.

It is interesting to note that the two meaning senses which were successfully recalled by the greatest number of participants (i.e. 119 out of 128) were *turn up* as in *turn up the radio*, and *come on* as in *come on, you can do it*. We can see that both of these meaning senses are figurative and yet were very widely known. In the case of *turn up*, we might suspect that many learners had been exposed to the item in the classroom (for example, when asking the teacher to increase the volume of an audio recording). In the case of *come on*, it might be that learners find the item particularly salient due to its prosodic and extra-lingual cues (typically, an exclamative tone and a hand gesture). In all cases, the effect of these few additional factors (saliency, context availability, phrasal patterning, etc) remains unclear, and could usefully be investigated in future research.

*4.2.2. Exposure factors: L2 instruction, L2 immersion and L2 engagement in leisure activities*

Our 110 participants were spread from 1st to 4th year of BA study, and had spent four years studying general English on average ($Min = 1$, $Max = 16$, $SD = 3$) prior to taking our test. As we can see, our final best-fit model includes neither year of BA study nor years spent studying English as predictors of knowledge. This may seem surprising, as we would expect that more L2 instruction results in more L2 vocabulary knowledge. In fact, we might suspect that the type of exposure students get from a general English course, and even a BA TEFL/English course, is not necessarily the sort of informal spoken discourse in which PVs can typically be found. Furthermore, multi-word combinations are typically neglected in favour of single words, as

**Table 3**
Main fixed effects on PV knowledge as identified by mixed models analysis.

| Predictor | Estimate | Std. Error | *z*-value | *p*-value |
|---|---|---|---|---|
| (Intercept) | −7.60096 | 1.33389 | −5.698 | 1.21e−08*** |
| Reading | 0.02720 | 0.01146 | 2.373 | 0.0177 * |
| Social Networking | 0.02774 | 0.01263 | 2.197 | 0.0280 * |
| Log COCA Frequency | 0.86142 | 0.17134 | 5.028 | 4.97e−07 *** |

Significance values are estimated by the R package lmerTest (version 2.0–11; Kuznetsova, Brockhoff, & Christensen, 2014): *** p < 0.001, ** p < 0.01, * p < 0.05.

vocabulary tends to be primarily conceived as individual words (e.g. single word entries in dictionaries). In spite of growing research interest in formulaic sequences, many teachers around the world are not aware of their importance and focus their teaching on single words only. Schmitt and Redwood (2011) also found that the type of instruction and hours of classroom input that their participants received did not have a significant effect on their scores.

Similarly, we found no effect of immersion in L2 country on knowledge. Our participants had spent one month abroad on average (*Min* = 0; *Max* = 18; *SD* = 3), which is arguably too little time to get the benefits of living in a country where the L2 is spoken. We may suspect that if participants had spent more time abroad overall, L2 immersion could have been a predictor of PV knowledge. However, Siyanova and Schmitt, 2007 found that the length of time spent in L2 environments did not have a strong effect on their participants' likelihood of using PVs instead of one-word verbs. As mentioned in our literature review, a more accurate predictor of formulaic language knowledge might be the degree of sociocultural adaptation, leading to a better quality of L2 exposure (Adolphs & Durow, 2004; Dörnyei et al., 2004).

Unlike L2 instruction and L2 immersion, the effect of L2 engagement in leisure activities on PV knowledge is clearly apparent from our results. In addition to frequency, our final model includes two predictors relating to everyday engagement: reading and social networking. This means that the more hours participants spent reading and social networking per week, the more knowledge they had. Conversely, watching films and listening to music did not have any effect on PV knowledge. These results are congruent with Schmitt and Redwood (2011) and González Fernández and Schmitt (2015) in that reading was identified as a predictor of knowledge whereas listening to music was not. But contrary to them, watching films did not have any effect in our study, and contrary to Author, we found social networking to be a predictor of PV knowledge in our study.

Reading has long been identified as a strong facilitator of vocabulary knowledge in the case of single words (Horst, Cobb, & Meara, 1998; Nation, 2001; Pigada & Schmitt, 2006), and so it is not surprising to find it as a main predictor in our model. We voluntarily included a wide spectrum of reading material under the 'reading' umbrella in our questionnaire (books, magazines, newspapers and even websites), and therefore it is likely that those who reported spending many hours reading per week get a type of L2 exposure that varies widely across genres, topics and registers. Such rich and varied exposure is typically conducive to acquiring a rich and diverse L2 vocabulary, among which multiword units such as PVs. As to social networking in the L2, we could explain its presence in our final model by the fact that, typically, using social media triggers the use of informal spoken language, and thus PVs. It is a type of exposure which makes language more engaging and personal, and thus perhaps even more conducive to learning.

Whilst watching films/TV is also conducive to varied and engaging exposure to everyday informal language, the L2 input is often in aural mode only and the subtitles, if any, are perhaps more commonly used in the L1. The lack of visual input (i.e. not being able to visualise the written form of words) may seriously impair learning and retention of L2 vocabulary. The same speculation can be made regarding listening to music, which in addition to being in aural mode, does not usually require much attention or concentration. This means that a great deal of unknown vocabulary might be left unnoticed as a result. Also, it is often hard to make out the words in songs, even for native speakers.

## 5. Conclusion

PVs are important for language use, but are widely considered challenging. So how much do learners know? It depends to some extent on how PV knowledge is measured. Previous studies have typically tested only one meaning sense (the most common one), but still found incomplete knowledge, e.g. Schmitt and Redwood (2011) at 48%. But when multiple meaning senses were tested in our study (taking into account the polysemous nature of PVs), the results were even lower. Our participants knew only about 40% of PV meaning senses on average, with only about a 20% chance that all the various meaning senses of each PV tested would be known. These are rather disappointing results considering that these learners were BA English/TEFL students in a partial English-medium academic environment, which lends further support to the common assertion that PVs are a problematic feature of English vocabulary for many learners, including presumably advanced ones. However, it is also important to note that our participants were tested at a relatively high *form recall* level of mastery, and use of a receptive test would probably have yielded much higher scores.

And what predicts knowledge? Semantic opacity did not seem to matter, but corpus frequency clearly won the battle. It was the best predictor of PV knowledge in our study, and this finding is consistent with previous research showing the robust effects of corpus frequency for both individual words and formulaic sequences. Among exposure-related factors, L2 instruction, L2 immersion, and year of BA study did not have any effect on knowledge. But L2 engagement in leisure activities clearly did: both reading and social networking seemed to promote the learning of PVs. This is good news as it suggests that it is possible to learn a lot outside the classroom, via daily activities that are engaging and enjoyable. Our study suggests that we should encourage L2 learners to spend more time doing such activities in order to increase their knowledge of PVs, and L2 vocabulary in general. However, if one wishes to effectively improve vocabulary knowledge, they might need to combine their implicit learning with explicit learning, and in the case of teachers, to give more attention to PVs in instructed contexts.

Finally, all of the findings above must be interpreted in light of the inevitable limitations of this study. The amount of L2 engagement of the participants was assessed via self-report questionnaires, and thus could be prone to slight underestimation or overestimation. The participants were a fairly homogeneous group (all BA English/TEFL students with the same L1), and therefore further research involving more diverse samples of L2 populations will be needed in order to make more robust generalizations. Investigating the effect of proficiency on knowledge would also be desirable.

## Acknowledgements

Our deepest gratitude goes to Daniel Muñoz Acevedo and Anthony Adams for recruiting our participants in their respective universities. We would also like to thank our reviewers for their insightful feedback on the study.

## Appendix A

## Appendix B

| PV | | Meaning sense | COCA frequency |
|---|---|---|---|
| Back up | F | Move or drive backwards a short way | 2974.66 |
| | F | Take action in order to support STH or make it happen | 2402.61 |
| | F | Establish as valid or genuine | 2345.405 |
| Break down | F | Stop working or functioning; fail or collapse (vehicle, device, relationship, negotiations) | 2106.72 |
| | F | Divide or separate into categories or smaller components so as to make it easier to understand or deal with | 1755.6 |
| | F | Lose control of one's emotions and yield to tears or distress | 1536.15 |
| | L | Undergo chemical decomposition; separate into different substances | 1185.03 |
| Break off | L | Separate a part (or become separate) from a larger piece | 843.6 |
| | F | Stop speaking, especially suddenly | 590.52 |
| | F | Put an end to STH (relationship, discussion, talks, negotiations) | 506.16 |
| Bring in | L | Bring STH to a place | 4248.92 |
| | F | Ask SB to do a particular job or task | 2492.155 |
| Clean up | L | Get rid of dirt or mess | 6147.92 |
| | F | Make STH free from dangerous, unacceptable or controversial activities or contents | 1827.76 |
| Come along | F | Appear or arrive; come into existence | 3969.375 |
| | L | Go somewhere with SB | 1122.375 |
| Come in | L | Enter a place or area (room, building) | 19765.2 |
| | F | Become involved in a situation | 4257.12 |
| Come on | F | Said to encourage SB to try harder, or do or say STH | 10445 |
| | F | Said to show SB disbelief, disagreement, or anger | 4073.55 |
| Come out | L | Leave a place (room, building, container) or appear from it | 13874.94 |
| | F | Become known or revealed after being kept secret | 4929.255 |
| | F | (+ *and do/say* …) Make public knowledge a privately held position | 4198.995 |
| | F | Become available or released to the public (film, book …) | 3651.3 |
| Cut off | L | Remove a part of STH by cutting it | 2036.88 |
| | F | Interrupt SB as they are speaking | 1848.28 |
| | F | End the provision of STH, or be deprived of a provision (electricity, money) | 1772.84 |
| Get down | F | (+ *to* …) Begin to pay serious attention to STH | 1957.8 |
| | L | Lower one's body as by kneeling, sitting or lying | 1694.25 |
| | L | Come down from STH; descend (car, horse, tree) | 1317.75 |
| Get off | L | Go away from, leave (train, bus, aircraft, lift) | 3884.22 |
| | F | (*Get off to a … start*) Begin STH in a certain way | 899.125 |
| | F | Manage to avoid serious trouble or consequences (esp. legal punishment) | 863.16 |
| Get on | F | (+ *with* …) Continue doing STH after stopping; proceed with STH | 2973.93 |
| | L | Board some form of public transportation (train, bus, plane, elevator) | 760.235 |

(*continued*)

| PV | | Meaning sense | COCA frequency |
|---|---|---|---|
| Give out | L | Give to each of a large number of people; distribute | 1209.6 |
| | F | Make known openly or publicly; reveal | 1013.04 |
| | F | Collapse, fail; stop functioning properly (heart, knees) | 347.76 |
| Go down | L | Move down to a lower level or position | 5827.84 |
| | F | Decrease in value or amount | 5425.92 |
| | F | Go from one place to another, esp. one that is further south or underneath | 3617.28 |
| Go up | F | Become higher in value; increase | 9424.95 |
| | L | Move upward, or from a lower spatial location to a higher one | 4067.61 |
| Hand over | L | Give STH to SB by holding it in one's hand and offering it to them | 1755.585 |
| | F | Surrender control or responsibility for STH/SB to SB else, esp. officially | 1245.415 |
| Hold back | F | Decide not to do or say STH | 988.645 |
| | F | Prevent SB/STH from reaching their full potential | 883.47 |
| | L | Prevent SB/STH from going somewhere | 736.225 |
| | F | Contain an unwanted physical manifestation (tears, laughter, sigh, sneeze) | 673.12 |
| Keep up | F | Move, progress or increase at the same rate or pace as SB/STH | 4897.16 |
| | F | Make STH continue | 3459.95 |
| Look back | F | Think of STH again, reconsider STH past | 7551.225 |
| | L | Look at STH/SB again after having momentarily looked elsewhere | 4576.5 |
| Look out | L | Look outside, or at the horizon | 6499.855 |
| | F | Take care of SB and make sure they are well; protect SB's interests | 3282.105 |
| Make out | F | See or hear with difficulty | 4052.29 |
| | F | Represent as being a particular way, esp. falsely | 736.78 |
| | F | (*Make it out*) Deal with a difficult situation successfully | 703.29 |
| Move up | F | Move to a better position; advance to a higher level/rank | 2101.37 |
| | L | Move upward, from a lower spatial location to a higher one | 1005.975 |
| Pay off | F | Pay the complete amount of STH | 3438.33 |
| | F | Pay back the effort spent in doing STH by becoming profitable or effective | 3403.245 |
| Pull back | L | Move backwards or make SB/STH move backwards | 3767.225 |
| | F | Withdraw or retreat from an activity or location, esp. military | 1756.15 |
| Put in | L | Put one thing inside another; include or insert | 1525 |
| | F | Invest or devote so as to achieve STH (time, effort, work) | 808.25 |
| Put on | L | Put a piece of clothing or jewellery onto one's body | 3341 |
| | F | Present or stage (play, show, competition) | 931.635 |
| Put out | F | Make STH known or accessible to the public (information, products) | 3899.59 |
| | F | Stop STH from burning or shining | 1161.58 |
| | L | Place STH somewhere in order for it to be seen or used | 829.7 |
| Put up | L | Display or attach STH (e.g. to a wall) so it can be seen | 2456.17 |
| | F | (+ *with* ...) Be willing to accept STH unpleasant or not desirable; tolerate | 2029.01 |
| | F | Build or place STH somewhere | 1922.22 |
| Reach out | L | Stretch an arm in order to hold, touch, or get STH that is within short distance | 4792.77 |
| | F | Make an effort to address or communicate with SB, so as to help them or involve them in STH | 3903.39 |
| Run out | F | Use STH (or become used) completely so that nothing is left | 2822.985 |
| | L | Leave suddenly, as if in a hurry | 1939.02 |
| Set out | F | Start doing or working on STH, esp. with a particular goal in mind | 3553.85 |
| | F | Start a journey | 2215.93 |
| | F | Explain or present STH clearly, esp. officially and in writing | 1337.92 |
| Sit back | L | Rest in a comfortable position against the back of a seat | 2671.68 |
| | F | Deliberately take no action/remain passive about STH | 1376.32 |
| Stand out | F | Distinguish oneself/itself by being better, more significant or more impressive than other people/things | 3318.425 |
| | F | Be easily seen or noticeable | 2084.3 |
| Take back | L | Take STH/SB to a place they were in before | 3224.5 |
| | F | Regain possession or control over STH | 2160.415 |
| Take in | F | Provide a place for SB to live or stay | 935.165 |
| | F | Fully understand or grasp the meaning of STH | 667.975 |
| | F | Deceive by behaving in a dishonest way | 381.7 |
| Take out | L | Remove or extract STH from a container | 7182.615 |
| | F | Invite to a recreational place or social event | 1920.105 |
| | F | Obtain an official document or service from an authority | 1777.875 |
| Turn around | L | Move so as to face in the opposite direction | 7696.35 |
| | F | Make STH become better or more successful than it previously was (economy, business) | 2793.49 |
| Turn over | F | Surrender possession or control to SB/STH (esp. in authority) | 3709.23 |
| | L | Change position so that the other side is facing towards the outside or the top, or another direction | 2119.56 |
| Turn up | F | Yield; be (or make STH be) found, discovered, or noticed | 3608.64 |
| | F | Increase the volume or level of STH | 1616.37 |
| | F | Arrive or make an appearance somewhere | 1052.52 |

L = Literal meaning sense.
F = Figurative meaning sense.

## Productive phrasal verb test

**Student:** _____

We are carrying out a study of students' productive knowledge of phrasal verbs. To help us in our research please complete this test.

Read each sentence carefully, and then write what you think the missing words (a phrasal verb) are, in the space next to the sentence. To help you, the first letter(s) of each word is/are shown. We have also given a definition for each phrasal verb after every sentence. **Please make sure you read each definition carefully**. There are 100 sentences and some of them use the same phrasal verb.

You have 30 minutes to finish the test. Good luck!

Example sentences:

| # | Sentence | Answer |
|---|---|---|
| i | The prisoners are hoping to g___ o___ of jail soon. (***leave***) | *get out* |
| ii | I didn't think he would b___ u___ the subject. (***mention, introduce***) | *bring up* |
| iii | She b___ u___ her children under very difficult circumstances. (***raised, educated***) | *brought up / bring up* |

| | | |
|---|---|---|
| 1 | Put the chicken on the grill and t___ it o___ a few times. (***bring the bottom to the top or vice versa***) | |
| 2 | You need to take the bus and g___ o___ at the third stop. (***leave the bus***) | |
| 3 | Their new album will c___ o___ next month. (***be released to the public***) | |
| 4 | She h___ b___ the laughter with great effort. (***contained, repressed***) | |
| 5 | Following a disagreement, the government br___ o___ diplomatic relations with China. (***ended***) | |
| 6 | They p___ u___ a few posters on the wall. (***displayed, attached***) | |
| 7 | After hitting the iceberg, the ship began to g___ d___. (***sink***) | |
| 8 | She p___ b___ the curtains so the light could come into the room. | |

| | | |
|---|---|---|
| | (***removed***) | |
| 9 | He t___ u___ to the meeting half an hour late. (***arrived, appeared***) | |
| 10 | He closed the dictionary and l___ b___ to his notes. (***watched again after watching something else***) | |
| 11 | This is amazing; k___ u___ the good work! (***continue***) | |
| 12 | They p___ o___ such an incredible show last night! (***presented, staged***) | |
| 13 | We s___ ou__ for San Francisco on the following day. (***left, departed***) | |
| 14 | We need experts to c___ i___ and give us advice. (***join, become involved***) | |
| 15 | After dinner, he t___ her b___ to her house. (***returned***) | |
| 16 | We l___ o___ for each other as if we were family. (***protect, take care of***) | |
| 17 | The storm c___ of___ electricity from the entire town. (***ended the provision of***) | |
| 18 | He had been b___ i___ to save the company. (*involved in a situation*, *introduced*) | |
| 19 | He was asked to cl___ u___ his language during the interview. (***make more acceptable/appropriate***) | |
| 20 | Unfortunately we've r___ o___ of biscuits. (***used completely***) | |
| 21 | We're going to the cinema tonight; you should c___ a___ with us! (***join***) | |
| 22 | Digestion b___ d___ substances into small molecules. (***decomposes***) | |
| 23 | We won't just si__ ba__ and watch the situation getting worse and worse. (***take no action***) | |
| 24 | I had to t___ o___ a loan to cover all my expenses. (***obtain***) | |
| 25 | You need to p___ i___ more hours at the office every day. (***invest, devote***) | |
| 26 | He got into his car and b___ u___ until he reached the street. (***drove backwards a short way, retreated***) | |
| 27 | The committee g___ o___ more than 100 copies in the last meeting. (***distributed***) | |
| 28 | She turned around to h___ o___ her keys to her husband. (***give, present***) | |
| 29 | We should g___ d___ to discussing those issues as soon as possible. | |

| | | |
|---|---|---|
| | (***begin***) | |
| 30 | Excellent product quality is what made the brand s___ o___ from its competitors. (***distinguish itself by being better***) | |
| 31 | C___ o___, don't be shy and tell us your story. (***said as encouragement***) | |
| 32 | Police have p___ o___ a warning about thieves in the area. (***issued, broadcast***) | |
| 33 | Oil prices have g___ u___ last year. (***increased***) | |
| 34 | She t___ ar___ and walked out the door. (***moved so as to face in the opposite direction***) | |
| 35 | The family t___ her i___ when she was abandoned by her parents. (***accommodated, sheltered***) | |
| 36 | It will take a dozen years for him to p___ o___ his debts. (***clear***) | |
| 37 | People need to c___ o___ and say what they think about it. (***declare publicly***) | |
| 38 | We might as well g___ o___ with it if we want to finish on time. (***proceed, continue***) | |
| 39 | The government has recently been trying to r___ o___ to right-wing voters. (***address, communicate with***) | |
| 40 | Security guards tried to h___ b___ the crowd. (***stop***) | |
| 41 | She se___ o___ to discover the truth behind the story. (***undertook, began with a definite purpose***) | |
| 42 | She m___ u___ from secretary to senior manager in just a few years. (***advanced, progressed***) | |
| 43 | I won't p___ u___ with your bad behaviour for much longer. (***tolerate***) | |
| 44 | I don't think prices will g___ d___. (***decrease***) | |
| 45 | Politicians often fail to b___ u___ their words with actions. (***support***) | |
| 46 | The team has g___ o___ to a good start this season. (***begun in a certain way***) | |
| 47 | He b___ d___ at his son's funeral. (***yielded to tears or distress***) | |
| 48 | G___ d___ on your knees so you can get a better view. (***lower body***) | |

| | | |
|---|---|---|
| 49 | I could barely m___ o___ his face in the dark. (***see***) | |
| 50 | I really like this song; could you t___ u___ the radio? (***increase volume***) | |
| 51 | The fire has finally been p___ o___. (***extinguished)*** | |
| 52 | We were lucky to m___ it o___ of the war alive. (***deal with a difficult situation successfully***) | |
| 53 | At 95 years of age, her heart finally g___ o___. (***collapsed, failed***) | |
| 54 | You cannot let a few unmotivated pupils h___ b___ the rest of the group. ***(limit potential***) | |
| 55 | He was telling me a story but b___ o___ abruptly when his mobile phone rang. (***stopped speaking***) | |
| 56 | I've p___ o___ glasses and a bottle of wine. (***displayed, made ready for use***) | |
| 57 | She liked to go by the window and l___ o___ at the garden. (***watch***) | |
| 58 | I didn't expect such an opportunity to c___ al___. (***appear, arrive***) | |
| 59 | Workers' income has not k___ u___ with inflation. (***increased as fast as***) | |
| 60 | You should t___ her o___ to this new Chinese restaurant. (***invite***) | |
| 61 | The government isn't willing to h___ o___ power to local authorities. (***surrender, yield***) | |
| 62 | She put her hand on his shoulder and m___ it u___ along the back of his neck. (***raised, lifted***) | |
| 63 | Let's b___ d___ the task into three easy steps. (***divide***) | |
| 64 | The teacher c___ o___ the student in the middle of her sentence. (***interrupted abruptly***) | |
| 65 | Flashing lights make planes s___ o___ at night. (***be easily seen or noticed***) | |
| 66 | He loves climbing trees but finds it hard to g___ d___. (***descend***) | |
| 67 | He could see a few hands g___ u___ in the audience. (***raise***) | |
| 68 | You have to explain more thoroughly; it's too difficult to t___ i___. (***understand***) | |
| 69 | The official recommendations were se___ o___ in the document. (***explained, described***) | |

| | | |
|---|---|---|
| 70 | They've p___ u___ a new fence after the previous one fell apart. (***built***) | |
| 71 | He g___ o___ the bus to school. (***boarded***) | |
| 72 | I b___ i___ my laptop computer today because my office computer is broken. (***took to a place***) | |
| 73 | She r___ o___ for the empty jar on the table. (***stretched an arm so as to grab***) | |
| 74 | The policeman tu___ o___ the criminal to the jail guard. (***transferred, surrendered***) | |
| 75 | Oh c___ o___, you're just lying to me! (***said to show anger or disbelief***) | |
| 76 | The politician's ultimate goal is to t___ b___ the Senate. (***regain possession of***) | |
| 77 | People have stopped believing the President could t___ a___ the economy. (***change dramatically for the better***) | |
| 78 | You have to b___ u___ your accusations with solid evidence. (***prove, establish as true***) | |
| 79 | The news c___ o___ that he was leaving the team. (***became known***) | |
| 80 | You should p___ o___ your gloves, it's really cold outside. (***wear***) | |
| 81 | They should not h___ b___ from joining us if they want to. (***refrain***) | |
| 82 | All the hard work will p___ o___ in the end. (***be worth it, reward***) | |
| 83 | He was innocent, but the media m___ him o___ to be a criminal. (***represented as being, especially falsely***) | |
| 84 | Make sure you c___ u___ your room because I won't do it for you. (***tidy***) | |
| 85 | The army was forced to p___ b___ due to bad weather. (***withdraw***) | |
| 86 | She opened the door and he c___ i___. (***entered***) | |
| 87 | You should be more careful and not g___ o___ your phone number so easily. (***reveal***) | |
| 88 | He was very convincing, so I was easily t___ i___. (***deceived***) | |
| 89 | He wants to g___ d___ to Australia next year. (***travel***) | |
| 90 | Take this application form and p___ i___ your name and contact details. (***include, insert***) | |

| | | |
|---|---|---|
| 91 | She thinks some criminals g___ o___ too easily. (***become cleared of a criminal charge***) | |
| 92 | She s___ b___ in her chair and turned on the TV. (***settled, rested***) | |
| 93 | Our car b___ d___ yesterday. (***stopped working***) | |
| 94 | Take the carrots and c___ the ends o___. (***remove***) | |
| 95 | Sometimes we l___ b___ on those days and realize we had a very happy life. (***think again, reconsider***) | |
| 96 | After the argument, she r___ o___ into the garden and screamed. (***left suddenly/in a hurry***) | |
| 97 | He tore open the envelope and t___ o___ a few bills. (***extracted, removed***) | |
| 98 | The search t___ u___ solid evidence against him. (***yielded, revealed***) | |
| 99 | He accidentally b___ o___ a piece of wood from the fence. (***removed, separated***) | |
| 100 | She went into the bank and c___ o___ with some money. (***exited, left***) | |

Thank you very much!

## Questionnaire

*In order to help us better understand and interpret your score, we would like to know a little bit about yourself and your experience as a language learner. Please provide answers to the following questions by ticking the boxes (□) or filling in the blanks.*

**Gender**: □ Male □ Female

**Age**: _____

**How many years have you been studying English?** _____ years

**How many months in total have you spent in English-speaking countries?**_____months

**How many hours per week do you spend:**

- **reading books, magazines and newspapers in English, or visiting English language websites?** _____ hours
- **watching films, videos or TV in English?** _____ hours
- **listening to music in English?** _____ hours
- **using English to keep in contact with people? (Facebook, MySpace, Twitter, Skype, email, SMS, etc.)**: _____ hours

*Finally, we would like to thank you very much for your participation. We really appreciate your help and contribution to this study. Thanks a lot!*

*If you would like to know more about the results of the study, please do not hesitate to contact us via email:* ***

## References


Adolphs, S., & Durow, V. (2004). Social-cultural integration and the development of formulaic sequences. In Author (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 107–126). Amsterdam: John Benjamins.

Annual Review of Applied Linguistics. (2012). *Topics in formulaic language* (Vol. 32). Cambridge: Cambridge University Press.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., et al. (2014). *lme4: Linear mixed-effects model using Eigen and S4, R package version 1.1-10*.

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

Chen, M. (2013). Overuse or underuse: a corpus study of English phrasal verb use by Chinese, British and American university students. *International Journal of Corpus Linguistics, 18*(3), 418–442.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than 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.

Dagut, M., & Laufer, B. (1985). Avoidance of phrasal verbs: a case for contrastive analysis. *Studies in Second Language Acquisition, 7*(1), 73–79.

Dirven, R. (2001). English phrasal verbs: theory and didactic application. In M. Putz, S. Niemeier, & R. Dirven (Eds.), *Applied cognitive linguistics II: Language pedagogy* (pp. 3–27). Berlin: Mouton de Gruyter.

Dörnyei, Z., Durow, V., & Zahran, K. (2004). Individual differences and their effects on formulaic sequence acquisition. In Author (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 87–106). Amsterdam: John Benjamins.

Ellis, N. C. (2002). Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*(2), 143–188.

Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: a corpus-based analysis. *TESOL Quarterly, 41*(2), 339–359.

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.

González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have?: The effects of frequency and amount of exposure. *ITL International Journal of Applied Linguistics, 166*, 94–126.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*(2), 207–223.

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

Kurtyka, A. (2001). Teaching English phrasal verbs: a cognitive approach. In M. Putz, S. Niemeier, & R. Dirven (Eds.), *Applied cognitive linguistics II: Language pedagogy* (pp. 29–54). Berlin: Mouton de Gruyter.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). *The R package lmerTest (version 2.0–11)*.

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.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English based on the British national corpus*. London: Longman.

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.

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics, 33*(3), 299–320.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P., & Waring, R. (1997). Vocabulary size, text coverage, and word lists. In Author (Ed.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6–19). Cambridge: Cambridge University Press.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: a case study. *Reading in a Foreign Language, 18*(1), 1–28.

R Development Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: the R Foundation for Statistical Computing, ISBN 3-900051-07-0. Available online at: http://www.R-project.org/.

Reppen, R., & Simpson-Vlach, R. (2010). Corpus linguistics. In Author (Ed.), *An introduction to applied linguistics* (pp. 89–105). London: Hodder Education.

Rudzka-Ostyn, B. (2003). *Word power: Phrasal verbs and compounds: A cognitive approach*. Berlin: Mouton de Gruyter.
Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, UK: Palgrave Macmillan.
Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: word associations and verbal suffixes. *Studies in Second Language Acquisition, 19*(1), 17–36.
Schmitt, N., & Redwood, S. (2011). Learner knowledge of phrasal verbs: a corpus-informed study. In F. Meunier, S. de Cock, G. Gilquin, & M. Paquot (Eds.), *A Taste for Corpora*. Amsterdam: John Benjamins.
Siyanova, A., & Schmitt, N. (2007). Native and nonnative use of multi-word vs one-word verbs. *International Review of Applied Linguistics, 45*, 109–139.
Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: a multi-study perspective. *Canadian Modern Language Review-Revue Canadienne Des Langues Vivantes, 64*(3), 429–458.
Waibel, B. (2007). *Phrasal verbs in learner English: A corpus-based study of German and Italian students*. Doctoral dissertation. Freiburg: Albert-Ludwigs-Universität Freiburg.
Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.