The PHRASal Expressions (PHRASE) List User's Guide

What is the PHRASE List?

The PHRASE List is a listing of the most common multiword expressions in English of a certain type, intended to be used as a complement to existing lists and instruments of second language instruction that use them.

What type of expression does the list contain?

There are of course many different types of multiword formulaic sequences in English (e.g. idioms, phrasal verbs, binomial expressions, collocations), and an exhaustive listing of those is not the aim of the PHRASE List. Instead, what the PHRASE List intends to provide is a full account of the most common expressions in English that also potentially could cause decoding problems if read word-for-word. For example, even beginners might know the verb *take* and the noun *place*, but may not know the phrasal expression *take place*, which of course has its own meaning.

How was the list made?

The list was created using the 100 million word British National Corpus. First, a full list of all recurring word combinations was extracted, and then the phrasal expressions were chosen from among those line by line and checked against selection criteria. The items selected were also confirmed by an external rater who showed the criteria could be applied consistently.

How can I interpret the frequency information in the list?

The PHRASE List is inclusive of phrasal expressions in the British National Corpus (BNC) to a frequency that matches the frequency of the top 5000 words, also from the BNC. This frequency threshold has been identified as having functional significance for single words (e.g. the ability to read and write at good levels of general proficiency), and the items in the PHRASE List provide one missing lexical element that previous wordlists generally had not accounted for.

Each line in the PHRASE List contains frequency information about each item (Figure 1). The first column, 'Integrated List Rank', shows where the item would rank in a list that integrated both single words and phrasal expressions. This information was included to give the user some perspective regarding just how common each expression is.

The column which immediately follows each phrase, entitled 'frequency', shows how many times that item recurs in the BNC. Each item has been lemmatised where appropriate, so the frequency of *take place*, for example, reflects the combined frequencies of *take place*, *takes place*, *took place*, *taking place* and *taken place*. Moreover, when a phrasal expression could also have a different, more transparent meaning (e.g. *I want to see the circus **when it comes to** town* vs. *My wife is a genius **when it comes to** numbers*), only the frequency of the less transparent meaning has been listed.

Finally, there are three columns that follow the 'Frequency' column: 'Spoken general', 'Written general', and 'Written academic' - reflecting three different discourse genres

identified by the authors as useful for a broad audience of English learners. At least one column of the genre frequency information contains three stars, indicating the genre in which that phrase is most common. If the phrase occurs equally commonly in more than one genre, then the same amount of stars are given (see 'rather than', Figure 1). An 'x' rather than a star is given to a genre in which the expression in question appears very rarely, or even not at all. In most cases, however, there were various levels of frequency, as in the expression 'so that' (Figure 1), which is occurs most commonly in general English conversation, less in written general-purpose English, and even less in academic writing.

Figure 1. Sample lines from the PHRASE List

Integrated List Rank	Phrase	Frequency (per 100 million)	Spoken general	Written general	Written academic	Example
631	RATHER THAN	21085	* * *	* * *	* * *	Children, **rather than** adults, tend to learn quickly.
635	SO THAT	20966	* * *	* *	*	Park it **so that** the wheels are curbed.

What are the practical applications of the list?

Just as there have traditionally been a number of applications for existing wordlists, the PHRASE List will have the same applications and potentially more. Vocabulary tests, for example, often use wordlists with the same type of frequency information that the PHRASE List uses in order to obtain a sampling of the test taker's breadth of vocabulary knowledge. The PHRASE List could also be applied for the same purpose, and even integrated into existing tests.

For language teaching practitioners (e.g. teachers, syllabus designers), the PHRASE List provides one means of prioritising and ensuring the inclusion of important items of vocabulary that perhaps have not been systematically integrated to language instruction previously.

Finally, there are plans already under way to develop a text analyser that would allow the user to download any text and check which phrases at the various frequency levels appear in it, which in turn can be used to help arrive at informed decisions regarding the relative difficulty of the text.

What are the limitations of the PHRASE List?

Like most tools, the PHRASE List should be used appropriately and selectively. First of all, the list is derived from a corpus, and no matter what the size, no corpus is 100% reflective of all language use, but is an artefact of the texts which it comprises and the aims of the compilers of the corpus. Hence, since the PHRASE List is derived from the BNC, and the BNC is reflective of mostly written British English only as modern as the early 1990s, clearly the list would be of very limited use for someone wanting to know which phrases are most common in spoken American casual conversation. However, as long as the user is mindful of these limitations and heeds the genre information provided in the list, the PHRASE List should prove useful for many English language teaching professionals and students alike.