

About Mid-frequency Readers

The Mid-frequency Readers are books within a controlled vocabulary which are available free for advanced learners of English as a foreign or second language. They can be used in extensive reading programs or for individual study and enjoyment. They are designed to help learners learn mid-frequency vocabulary and to read texts that would otherwise be too difficult for them.

Published series of graded readers end at around the 3,000 to 4,000 word family level, but learners need a vocabulary size of 8,000 to 9,000 word families to read novels and newspapers. The Mid-frequency Readers are designed to provide interesting, comprehensible reading to fill this gap of 6,000-7,000 words between the end of graded readers and the demands of unsimplified text. The following references provide more information about mid-frequency words.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-82.

Schmitt, N. and Schmitt, D. (2012). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*

The Mid-frequency Readers are available free, and can be used for any purpose without permission, as long as they are not offered for sale, or are offered on web sites where payment must be made for access to the site. The adaptation is done in the same spirit that was behind the setting up of Project Gutenberg. It has been done without payment and without the wish for financial profit.

Ideally, each book should be read electronically on a tablet, e-reader or computer. This would allow easy electronic look up of unknown words. If the book is read in Microsoft Word, right click on unknown words to access Look_up, Synonyms, or Translate.

It is also possible to read the book accompanied by a spoken version using a text-to-speech program such as Dragon Naturally Speaking.

Each Mid-frequency Reader is in a controlled vocabulary and is adapted from the Project Gutenberg version of the text. Most of the words beyond a specified 1000 word level have been removed, largely through replacement by higher frequency words, but very occasionally adjectives and adverbs and the occasional short sentence have simply been omitted where this does not affect the story line. Where misspelling is used to represent accented speech (orl = all, nuff = enough), this has been changed to regular spelling. No other changes have been made to the book.

The aim of the changes is to make the books more accessible for non-native speaking learners of English by removing the large number of low frequency words which are way beyond their vocabulary level. The goal is to adapt the book so that in each book much less than 2% of the running words (tokens) are unfamiliar words, and there are no more than two to four hundred unknown different word families, depending on the length of the text. These unknown words can be dealt with by guessing from context, looking them up, or ignoring them.

Each book is adapted to three different levels.

1 The 8,000 level is for learners with a vocabulary size of 8,000 words. The words at the 9th 1000 and 10th 1000 levels are the target words, and the words beyond the 10th 1000 level have been replaced.

2 The 6,000 level is for learners with a vocabulary size of 6,000 words. The words at the 7th 1000 and 8th 1000 levels are the target words, and the words beyond the 8th 1000 level have been replaced.

3 The 4,000 level is for learners with a vocabulary size of 4,000 words. The words at the 5th 1000 level are the target words, and the words beyond the 5th 1000 level have been replaced.

Most words beyond the target level which occur four times or more in a particular book are not removed. The repetition of these words provides a good chance that they might be learned. A few words that are important for the message of the text and are very difficult to replace can be left in the text.

To find your vocabulary size so that you can choose books at the appropriate level, sit the Vocabulary Size Test at www.lex tutor.ca or <http://my.vocabularysize.com/select/test>.

The word lists used for the adaptation are based on the BNC and COCA and currently go up to the 25th 1000 level. They are accompanied by lists of proper nouns, marginal words (gee, ow, um), transparent compounds and abbreviations. The AntWordProfiler program developed by Laurence Anthony has been used in the adaptations. We are grateful to Compass Media in Korea for support in developing this program.

Paul Nation