Core Vocabulary

Elfrieda H. Hiebert

What is the core vocabulary?

What texts support students' proficiency with the core vocabulary?

Should students be taught the core vocabulary as sight words?

Can the words in the core vocabulary be taught?

Overview

Success in the digital age depends on comprehending complex text. Within any complex text, the majority of words come from a very small group of words in written English. On average, 90% of the words in a text are drawn from 4,000 simple word families (e.g., help, helping, helps, helped, helper but not helpless or helpful). The other 10% of the words in texts come from the remaining 300,000 (or more) words in the English language. Many words in this latter group don't occur often (less than once per million) but, as extended vocabulary, they give text much of its distinctiveness. In stories, the rarer words enhance interest and understanding of characters, setting, and plot. In informational texts, the rarer words are often technical terms that give precision and clarity to the content. The extended vocabulary is often what is new for readers and increases the complexity of a text. When readers aren't adept with the core vocabulary (i.e., the 4000 simple word families), they have few resources to deal with the new and unique vocabulary.

The core vocabulary may appear to be simpler than the extended vocabulary. That is not necessarily the case. Very complex texts can be written with the simplest of words from the core vocabulary (e.g., *To be or not to be*). Many words belong to the core vocabulary because of their multiple meanings and uses. For example, words such as *set* and *back* have many meanings and also take on different parts of speech.

Educators' concerns are addressed in this issue of *TextProject Answers.*

What is the core vocabulary?

A rich vocabulary is essential to success with complex texts. English has an enormous number of words—approximately 300,000 active words. But all of these words don't carry the same weight.

One analysis of K–12 texts from all subject areas showed that about 5,600 words accounted for 80% of the words in texts. An additional 150,000 words accounted for the other 20% of the words.

When the 5,600 words are organized as word families—such as help, helps, helped, helping, helper—the number becomes smaller—4,000 simple word families. And when family members that don't occur often are included, these 4,000 simple word families account for 90% of the words in texts—including the texts of high school and college.

This group of 4,000 simple word families forms the core vocabulary.

The core vocabulary includes words for many concepts (e.g., *length*) as well as highly frequent words (e.g., *the*) and general academic words (*compare*). Facility with the core vocabulary is fundamental to reading success.

A list of the 4,000 simple word families can be downloaded at: http://www.textproject.org/library/resources/wordzones-for-4000-simple-word-families/
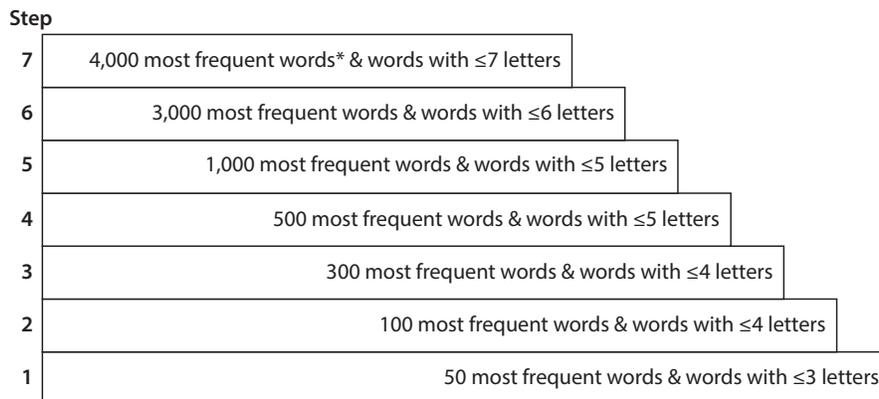
What is the difference between accessible and dumb-down text?

Providing engaging texts with an abundance of the core vocabulary is the raison d'être of TextProject. Suggestions for texts follow but educators need to remember that students need an enormous number of texts to become proficient readers (especially when the goal is reading complex texts). Educators often forget the number of book experiences of preschoolers in their acquaintance. Expecting children whose reading experiences happen mostly in school to become adept readers with one or two books per week is unreasonable. No one knows for certain but a reasonable hypothesis is that reading at least one million words by fourth grade underlies proficient reading.

TextProject has provided free, prototype programs of engaging texts with high percentages of core vocabulary. These programs have been created to serve as templates for selecting programs. Even with 120 *BeginningReads*, educators will need many more texts to teach children to read well. There are likely at 50,000-10,000 titles within the reading instructional market just for beginning reading in the American instructional market. At the present time, TextProject, as a small nonprofit, does not have the financial resources to independently analyze the many books and programs in the marketplace. But what TextProject does do is to provide exemplar texts which educators can use with students and, when their students experience success, to use these exemplar texts to locate additional texts among the thousands of texts available in the educational marketplace.

The staircase of core vocabulary (figure 1) is the basis for the three prototype programs: *BeginningReads*, *SummerReads*, and *Talking Points for Kids.*

Figure 1
Staircase of Core Vocabulary

Step

Step	
7	4,000 most frequent words* & words with ≤7 letters
6	3,000 most frequent words & words with ≤6 letters
5	1,000 most frequent words & words with ≤5 letters
4	500 most frequent words & words with ≤5 letters
3	300 most frequent words & words with ≤4 letters
2	100 most frequent words & words with ≤4 letters
1	50 most frequent words & words with ≤3 letters

* Includes simple word family members (e.g. *help, helping, helps, helped, helper*)

Table 1
TextProject Programs & the Staircase of Core Vocabulary

Program	Support for Core Vocabulary
BeginningReads	Across entire program, BeginningReads move students from Step 1 through Step 5 of the Staircase of Core Vocabulary.
SummerReads	Intended for middle-graders, the three levels focus on Step 5 of the Staircase (Grade 3), Step 6 (Grade 4), and Step 7 (Grade 5).
Talking Points for Kids	Intended for middle graders, the program emphasizes Step 7

Should students be taught the core vocabulary as sight words?

Recognizing words in the core vocabulary automatically is clearly a foundation of proficient reading. Because these words are so important, there can be a temptation to think that the core vocabulary should be taught through rote activities such as word lists or flashcards.

These activities are misdirected. The core vocabulary is learned through many, varied reading and writing experiences. Just because words are highly frequent does not mean that they are simple. Varied and numerous experiences are needed because most of the words in the core vocabulary take on many roles and meanings in written English.

"Many Uses: The 100 Most-Frequent Words in Written English" (available for download at: http://www.textproject.org/library/resources/many-uses-the-100-most-frequent-words-in-written-english/) illustrates the complexity of even the most-frequent words. Many of these words can be confused with other words because they are homophones (words that sound the same but have different spellings and meanings). Even a word with a single spelling can have different forms with unique origins and meanings (e.g., down, just) and many words with

a single origin are used as different parts of speech (*call, may*). Many common words also are used in idioms (phrases that don't have a literal meaning) and in compound words where the root words help with the meaning but where the parts don't automatically equal the whole (e.g., two-faced does not mean that someone literally has two faces).
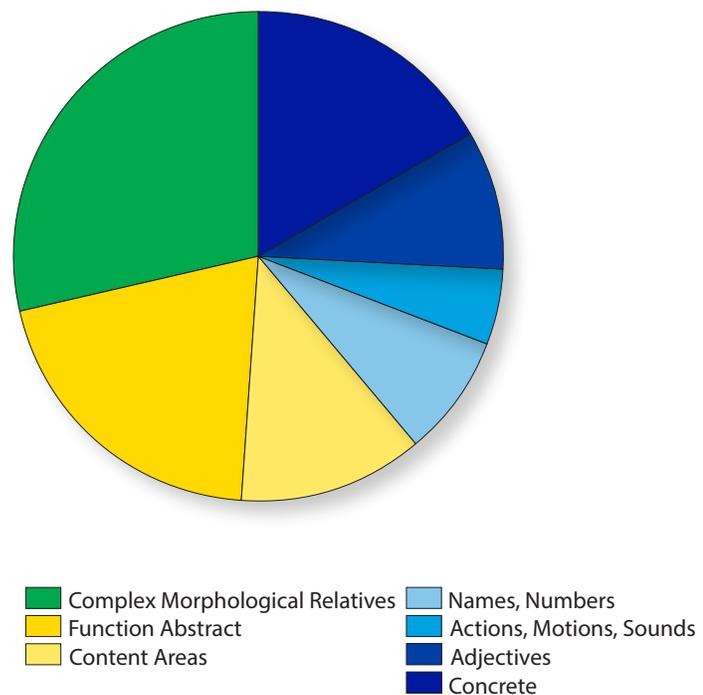
Becoming flexible with the meanings and uses of the core vocabulary requires numerous opportunities to read, write, and discuss texts. TextProject.org has a feature called *TextProject Word Lists* (accessible at: http://www.textproject.org/teachers/word-lists/). These word lists are intended for educators to identify vocabulary to teach, not for students to memorize in or out of school.

Can the words in the core vocabulary be taught?

The answer to the question is an unequivocal yes. But instruction does not take the form of rote activities. Students need many experiences in using the words in reading, writing, and discussion. These experiences include vocabulary lessons on different types of words within the core vocabulary.

The core vocabulary provides substantial grist for vocabulary lessons, due to the considerable variety among these words. Students benefit from instruction that brings the features of particular groups of words to their attention. The chart below illustrates the variety within the core vocabulary.

Figure 2
Distribution of the 4,000 Simple Word Families

- ■ Complex Morphological Relatives
- ■ Function Abstract
- ■ Content Areas
- ■ Names, Numbers
- ■ Actions, Motions, Sounds
- ■ Adjectives
- ■ Concrete

There are basically three types of words within the core vocabulary: (a) words that are often both familiar and concrete (sections in shades of blue); (b) abstract words (sections in shades of yellow); and (c) complex morphological relatives of root words (the green section).

Each type of word (familiar and concrete, abstract, and complex morphological relatives) has different features which should be the focus of lessons.

▶ Familiar and Concrete Words

The list of animal words is available in Word Lists in Teacher Resources on Text-Project. As the words on this list illustrate—goat, pony, hen, chicken, duck, pig—most are highly concrete and can be taught in lessons which make heavy use of pictures (see, e.g., sets of word pictures for vocabulary clusters at TextProject: Word Pictures.

▶ Complex Morphological Relatives

By applying a handful of guidelines, the 4,000 simple word families can be condensed even more—by about a third. Here are the guidelines:

• compound words with the "head" (first) word (example: night—midnight, overnight, tonight)
• words with the prefixes "a" and "un" with the root word (example: sleep—asleep; happy—unhappy)
• words with the suffixes "ment," "ion", and "ance" with the root word (example: arrange—arrangement; appear—appearance).

▶ Abstract Words

The abstract words within the core vocabulary are of three types:

• Dolch Plus: Function words (prepositions, pronouns, determiners, conjunctions, auxiliary verbs, particles)
• Academic Word List (AWL): Based on frequency of use in college texts, this list contains 570 words used to communicate processes and ideas in complex texts (Coxhead, 2001). Sixty percent of the AWL words (343 words) are in the 4,000 simple word families. These 343 words that appear on both Coxhead's list and on the 4,000 simple word families are listed here.
• Core Academic Word List (CAWL): The CAWL is based on frequency in elementary and high school texts and is limited to words within the 4,000 simple word families (but excluding the Dolch Plus and the AWL).

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.