DOCUMENT RESUME

ED 314 711   CS 009 856

AUTHOR   Marzano, Robert J.; Marzano, Jana S.
TITLE   A Study of Basic Words in Elementary School
                Textbooks.
INSTITUTION   Mid-Continent Regional Educational Lab., Aurora,
                CO.
SPONS AGENCY   Office of Educational Research and Improvement (ED),
                Washington, DC.
PUB DATE   May 89
NOTE   14p.
PUB TYPE   Reports - Research/Technical (143)

EDRS PRICE   MF01/PC01 Plus Postage.
DESCRIPTORS   *Basic Vocabulary; Elementary Education; Reading
                Research; Reading Strategies; *Textbook Content;
                Textbook Research; *Word Frequency; Word Lists
IDENTIFIERS   Prefixes; Suffixes; *Words

ABSTRACT
        A study was conducted to answer the question, "What
are the basic words in a list of words kindergarten through sixth
grade students will frequently encounter in their reading?" The study
was limited to the words the students would encounter in their
textbooks (including basals). A corpus of 7,230 words was then
analyzed by two raters to determine those that were basic. The major
finding was that out of 7,230 high frequency words, 5,084 were found
to be basic. The study also indicated that there are identifiable
high frequency basic words that can be used as a tool for classroom
instruction. Specifically, the 5,084 words identified in the study
appear quite frequently in the content-related materials students
will commonly encounter in their reading. The study of suffixes
indicates that instruction in their use might be most effective if it
focused on those that change basic words to noun forms. It would also
appear useful to provide students with an awareness of the dynamics
of change along the concrete/abstract and specific/general
dimensions. (Three tables of data are included.) (MG)

A STUDY OF BASIC WORDS
IN ELEMENTARY SCHOOL TEXTBOOKS

by

Robert J. Marzano

and

Jana S. Marzano

Mid-continent Regional Educational Laboratory

May, 1989

The concept of basic words has been of interest to educators for years, mainly for reasons of their assumed utility. Given that basic words are considered to be those from which other words are derived, a reasonable assumption has been that a knowledge of the basic words in the English language will naturally lead to increased facility at learning the other words in the language. Specifically, it has been assumed that teaching students a small set of basic words would circumvent the problem of the sheer numbers involved in teaching students all words that they will encounter in their reading (Becker, Dixon & Anderson-Inman, 1980).

One of the earliest attempts to identify the basic words in English was Ogden's (1932) "basic English," an 850-word lexicon from which all other English words could be derived. Unfortunately, Ogden's basic words were more like "primitive" concepts that could be used to map the semantic features of words, than they were like basic words from which other derived forms could be induced. Later, Dupuy (1974) operationally defined a basic word as one that is included in four major dictionaries, is not compound or hyphenated, a proper name, an abbreviation, foreign, archaic, informal, technical, derived or variant. Based on a one percent sample from the four criterion dictionaries, he then estimated that there are 12,000 basic words in the English language.

Following the Dupuy criteria, Becker et al. (1980) then identified 8,109 basic words from a list of 25,782 words drawn from an updated version of the Thorndike & Lorge (1941) list. The intent of their study was to create a vocabulary list that could be used as an instructional tool. Presumably, a knowledge of the 8,109 words on their list would allow one to infer the meaning of the 25,782 (and perhaps more) words from which the list was derived. However, Nagy and Anderson (1984) noted that Becker et al.'s use of a morphological basis for identifying basic words rendered their list impractical for educational

purposes. Specifically, although many words are related at a morphological level, they were not be related closely enough semantically for the average language user to make a connection. For example, *animism* and *animosity* were assigned the basic word *anima* by Becker et al. Nagy and Anderson (1984) implied that the vast majority of students would probably not know *anima* and, consequently, could not use it to understand *animosity*.

The most recent study of basic words was that done by Nagy and Anderson (1984). Using a corpus of written school English compiled by Carroll, Davies and Richman (1971), Nagy and Anderson estimated that there are 88,500 basic words (that they refer to as "word families")—about seven times that estimated by Dupuy (1974) for grades K through 12. Based on their estimate, Nagy and Anderson asserted that any attempt to identify and subsequently teach basic words to K through 12 students would be futile by virtue of the sheer numbers involved.

Beck, McKeown and Omanson (1987), however, offered an alternative to the Nagy and Anderson position. They explained that effective basic vocabulary instruction does not have to include all words students will encounter in their reading, only those that are high frequency words and/or are important to the understanding of specific content. They estimated that about half of the 88,500 word families calculated by Nagy and Anderson would be encountered only about once in an avid reader's lifetime. Of the remaining 44,250, only about 15,000 would be encountered once or more in 10 million running words. In short, Beck et al. conceded the impossibility of teaching all basic words students will encounter but argued for teaching the basic words identified from a corpus of relatively high frequency words.

Given the validity of the Beck et al. position, it would seem a useful endeavor to identify the basic words from a corpus of words students will commonly encounter in their reading. Consequently, this study sought to answer the question,

"What are the basic words in a list of words K through 6 students will frequently encounter in their reading?"

METHOD

Corpus

The intent in selecting an initial corpus for study was to identify or construct a set of words students frequently encounter from which basic words could be identified. Beck et al. (1987) estimated that there are 15,000 such words in grades 3 through 9. However, their estimate was based on the 86,741 word corpus of Carroll, Davies and Richman (1971) that was drawn from a wide variety of reading material students might encounter. That material included poetry, novels and general nonfiction. Although these certainly represent types of materials students might read, a more restricted corpus was identified for the present study. Specifically, it was decided to limit the corpus to words students will encounter in their textbooks (including basals). It was also decided to limit the study to grades K through 6. This was done under the assumption that the words found in K through 6 content materials are, for the most part, the higher frequency words in the English language.

The initial corpus selected for study was the Basic Elementary Reading Vocabulary (Harris & Jacobson, 1972). This text is based on 14 elementary school series (six basal series and two series each in the fields of English, social studies, math and science). The Harris & Jacobson corpus includes 7,613 words. Because it is somewhat dated, the list was reviewed by 60 elementary school teachers who: (1) deleted words that were not frequently encountered in instructional situations, and (2) added words that were commonly the focus of instruction. For addition of a new word in the corpus, a majority of the 60 teachers had to agree; similarly, a majority of the 60 teachers had to agree for a word to be excluded from the list.

The addition and deletion of words from the corpus resulted in a final list of 7,230 words.

Procedure

The corpus of 7,230 words was then analyzed by two raters to determine those that were basic. The following criteria were used in the identification of basic words:

1. Masculine forms of words were considered basic in cases where there were masculine and feminine forms.

Basic	Derived
duke	duchess
prince	princess

2. Singular forms of words were considered basic.

Basic	Derived
man	men
child	children

3. Neuter or androgynous forms of words were considered basic in cases in which there were neuter, masculine and feminine forms.

Basic	Derived
cowhand	cowboy
	cowgirl

4. Foreign words were considered basic (e.g., kimono).

5. Fields of study were considered basic, whereas names of practitioners within a field were considered derived.

Basic	Derived
science	scientist

6. Cardinal numerals were considered basic, whereas ordinal numerals were considered derived.

Basic	Derived
ten	tenth

7. Names of places were considered basic (e.g., Seattle, Canada).

8. Phrases were considered basic (e.g., kind of, because of).

9. Mature forms of living things were considered basic, whereas forms
   indicating developmental stages were considered derived.

Basic	Derived
chicken	chick

10. Technical terms were considered basic (e.g., cerebrum, cerebellum).

11. All pronoun forms were considered basic (e.g., everyone, someone).

12. Uncontracted forms of contractions were considered basic.

Basic	Derived
cannot	can't

13. Infinitive forms of verbs were considered basic, whereas other forms were
    considered derived.

Basic	Derived
do	doing

14. All directions were considered basic (e.g., northwest, southeast).

15. Species types were considered basic (e.g., rattlesnake, redwood).

16. Words indicating dimensionality or location were considered basic (e.g.,
    widespread).

17. Compound words were considered basic (e.g., horseback, snowplow).

18. Words formed by the addition of affixes to a root word were considered
    derived (e.g., unhappy).

19. Words that were not semantically related to any other word were considered
    basic.

20. Words indicating time were considered basic (e.g., noontime).

To some extent, these rules paralleled those established by Nagy and Anderson
(1984). Nagy and Anderson defined basic words as those that are semantically

opaque as opposed to semantically transparent. Semantically transparent words are those whose meaning can be derived by a knowledge of some root plus an affix or inflectional ending or, in the case of compound words, the meaning can be derived from a knowledge of the component words. Semantically opaque words are those that might be related to some other more basic words morphologically, etymologically, or even semantically, but that relationship is so weak (opaque) that the word could not be inferred by the average reader. Basic words, according to Nagy and Anderson, also include those that are not related etymologically, morphologically, or semantically to another word.

In effect, Rules 1, 2, 4, 5, 10, 12, 13, 18 and 19 operationalize Nagy and Anderson's notions of semantic transparency versus opaqueness. Rules 3, 6, 7, 8, 9, 11, 14, 15, 16, 17 and 20 are either substantially different from the Nagy and Anderson criteria or are not covered by their criteria. For example, Rules 3 and 8 were not specifically covered by their criteria. However, one can infer from the description of their study that they probably utilized similar rules. The other rules (6, 7, 9, 11, 14, 15, 16, 17 and 20) appear contradictory to their criteria. For example, numerals (Rule 6), proper names (Rule 7) and phrases (Rule 9) were not considered in their corpus and, thus, could not be counted as basic in their analysis. Rules 11, 14, 15, 16, 17 and 20, in the present study, automatically designated certain words as basic, where Nagy and Anderson made word-by-word decisions as to the semantic transparency versus opaqueness of words covered by these rules. To illustrate, Nagy and Anderson analyzed all compound words. In this study, compounds covered by rules 11, 14, 15, 16 and 20 were considered basic because of the perceived uniqueness of words in these categories. That is, the raters judged words designating pronouns (e.g., everyone) as basic because it was determined that such words play a central function in the English language and should, therefore, receive instructional attention (Rule 11). The same reasoning

applied to directions (Rule 14), species types (Rule 15), dimension/direction words (Rule 16), and words indicating time (Rule 20). All other compounds were considered basic (Rule 17) primarily because of the low inter-rater reliability within the study. Specifically, the inter-rater reliability for compound words not covered by Rules 11, 14, 15, 16 and 20 was .47 for the two raters. Therefore, it was concluded that compound words not covered by Rules 11, 14, 15, 16 and 20 represent a unique class of words and should therefore be considered basic.

RESULTS

The analysis of the 7,230-word corpus is reported in Table 1.

Table 1

	N
Corpus	7,230
Basic Words	5,084
Derived	2,254

As Table 1 indicates, 5,084 words of the 7,230-word corpus were identified as basic, and 2,254 were identified as derived. (The basic and derived words in their entirety are reported in Marzano and Marzano, 1988.)

Additionally, 41 types of prefixes and 77 types of suffixes that transform basic words to derived words were identified. The suffixes were further analyzed in terms of the syntactic change they effected in the basic words to which they were added. (Prefixes were not analyzed because they effect no change in the syntactic form of the words to which they are applied.) Table 2 reports the results of this analysis.

Table 2

Number of Suffixes Changing Basic Words of Specific Syntactic Forms to

Derived Words of Specific Syntactic Forms

Form of Derived Word

		N	Adj	Adv	V	TOTAL
Form of Basic Word	N	17	12		3	32
	Adj	15		1		16
	Adv				1	1
	V	22	5		1	28
	TOTAL	54	17	1	5	77

Table 2 indicates that the general direction of syntactic change was to nominal forms. That is, suffixes added to basic words that were adjectives, adverbs, verbs, and even other nouns commonly produced nominal forms for derived words. For example, 17 of the 77 types of suffixes identified were added to basic words that were nouns and transformed them into derived words that were also noun forms. Fifteen of the 77 types of suffixes were added to basic words that were adjectives and transformed them to derived words that were nouns, and so on. In all, 54 of the 77 suffixes that were added to basic forms generated derived words that were nouns.

The 77 suffixes were also analyzed in terms of the semantic changes they generated. The results of this analysis are reported in Table 3.

Table 3

Type of Semantic Change	N
Makes general or abstract something specific or concrete	25
Makes specific or concrete something general or abstract	25
Changes case between objects, agents, instruments, benefactors	20
Changes degree or order	7

Table 3 indicates that semantic changes in derived words created by the addition of suffixes were evenly distributed over three types: (1) specific/concrete to general/abstract, (2) general/abstract to specific/concrete, and (3) change from one case to another (e.g., agent to object, instrument to agent, and so on).

If one considers 1 and 2 above as changes of the same type with different directions, then Table 3 indicates that about 65% of the changes (50 of 77) were ones involving the concrete/abstract and specific/general dimensions.

Discussion

The major finding in this study was that out of 7,230 high frequency words, 5,084 were found to be basic. This supports the Nagy and Anderson (1984) assertion that there are far more basic words in English than originally estimated by Dupuy (1974). Even under Nagy and Anderson's assumption that the probability of a word being basic has a relatively high correlation with the frequency of the word (i.e., there are more basic words that have a high frequency of occurrence than there are basic words with a low frequency of occurrence), this

study indicates that the majority of words a student would encounter in school-related material are basic.

However, this study also indicates that there are identifiable high frequency basic words that can be used as a tool for classroom instruction. Specifically, the 5,084 words identified in this study appear quite frequently in the content-related materials students will commonly encounter in their reading.

Direct instruction in 5,084 basic words over an extended period of time (e.g., throughout grades K through 6) is not an insurmountable task. Given that these words occur with high frequency and that by virtue of the fact that they are basic they provide access to other words, they would seem to be strong candidates for direct instruction. This is not to imply that all basic words should be taught, nor that all students should receive direct instruction. Specifically, given that developmental readers learn the vast majority of vocabulary words incidentally from wide reading and that many of the high frequency words are learned incidentally even by poorer readers (Nagy, 1988), one might conclude that: (1) only students who are having difficulty with their reading development should receive direct instruction in the basic words identified here, and (2) those students should receive direct instruction only on those words that they have not already learned incidentally. Ideally, then, poorer readers could be quickly screened as to which words in the 5,084 corpus they were familiar with and receive instruction on those with which they were not familiar. A strategy for teaching these unknown basic words using semantic categories has been described by Marzano and Marzano (1988).

Additionally, the study of suffixes indicates that instruction in their use might be most effective if it focuses on those that change basic words to noun forms. In effect, most concepts tended to be nominalized by the suffixes identified in this study. An awareness, then, on the part of students of the characteristics of

nominalization might increase their understanding of many words they encounter. From a semantic perspective, it would also appear useful to provide students with an awareness of the dynamics of change along the concrete/abstract and specific/general dimensions. That is, students' understanding of new words they encounter might be facilitated if students grasped the dynamics of semantic changes affected by suffixes on the abstract to concrete continuum and on the general to specific continuum.

References

Beck, I., McKeown, M., & Omanson, R. (1987). The effects and uses of diverse vocabulary instructional techniques. In M. McKeown & M. Curtis (Eds.), The nature of vocabulary acquisition. Hillsdale, NJ: Erlbaum.

Becker, W., Dixon, R., & Anderson-Inman, L. (1980). Morphographic and root word analysis of 24,000 high frequency words (Tech. Rep. 1980-1). Eugene, OR: University of Oregon, Follow Through Project.

Carroll, J. B., Davies, P. & Richman, B. (1971). The American Heritage word frequency book. Boston: Houghton-Mifflin.

Dupuy, H. P. (1974). The rationale, development and standardization of a basic word vocabulary test (DHEW Publication No. HRA 74-1334). Washington, DC: U. S. Government Printing Office.

Harris, A. J. & Jacobson, M. D. (1972). Basic elementary reading vocabulary. New York: Macmillan.

Marzano, R. & Marzano, J. (1988). A cluster approach to elementary vocabulary instruction. Newark, DE: International Reading Association.

Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? Reading Research Quarterly, 19, 303-330.

Nagy, W. E. (1988). Teaching vocabulary to improve reading comprehension. Urbana, IL: National Council of Teachers of English and the International Reading Association.

Ogden, K. C. (1932). The basic words: A detailed account of uses. London: Kegan Paul.

Thorndike, E. L. & Lorge, I. (1941). The teacher's word book of 30,000 words. New York: Teacher's College Press.