4-1-1976

Some Semantics of Basic Word Lists

Patrick Groff
San Diego State University

WESTERN MICHIGAN UNIVERSITY

SOME SEMANTICS OF
BASIC WORD LISTS

Patrick Groff
SAN DIEGO STATE UNIVERSITY

Lists of words compiled on the basis of their frequency of occurrence in writings of various kinds have long been considered "basic" materials for reading instruction. It is said teachers should use such lists because the child best learns to read if the order of the words presented to him for this purpose is governed by the relative frequency words occur in written materials. Put another way, it is argued that if word A occurs more often in written matter of different types than does word B it then should be presented for a child to learn to read before word B is presented. There is a certain logic to support this procedure. Unknown words that appear frequently in a child's reading material surely pose more of a handicap to his success here than would unknown words that appear infrequently. Consequently, one can agree with Hillerich[1] that "most educators accept the need for a basic list."

The lists of basic words contrived on the basis of the frequency of their appearance have been numerous. Hillerich[1] refers to fourteen of these lists. I noted six others beyond the ones he identified.[2] Some of these basic word lists are not developed from newly uncovered evidence, but rather from lists that have been previously published. It is held by some that if a word is common to several basic word lists it is more basic than if it were found only in a single list.

Thus, Hillerich[1] contends that his list of basic words, which is constructed from several previous lists, is better than any of the single lists he consulted. This is so, he says, because it is "an updated basic vocabulary that minimizes the bias of individual counts." It is updated, he claims, since it does not include the "rural and childlike words little used today" Hillerich found in prior lists, and because it includes certain modern "technological words." The "bias" of any previous list is held to a minimum in his list, Hillerich goes on, for it is based on word counts which "include the old and new, juvenile and adult writing, and juvenile and adult printed material."

The title Hillerich gives for his explanation of the method he used to compile his derived word list, "Word Lists Getting It All Together,"[1] implies he believes the decisions he made here are all one needs to make about this matter. Hillerich overlooked one critical aspect of linguistics, however, that should be involved when any word list for teaching reading is put together. This is a consideration of the semantics, or meaning components, of these words. The compilers of word lists so far,[3] including Hillerich, generally have not considered the semantics of the words they studied. That is, in collecting these words they ignored the problems of how the varied number of potential denotations of such words might affect a

child's ability to read them. Lorge's semantic word count[4] is the notable exception to this.

The experts in language who study semantics are quick to remind us, however, that a meaning of a word usually cannot be determined when it is seen in isolation (as is the case in word lists). To understand the peculiar denotation or connotation given a word means a reader must depend on the context in which the word has been placed. Good writers are usually aware of the problems faced by the reader as he tries to get to the author's meaning. We find writers frequently will deliberately repeat a given word, often using a second sentence of a different structure for this. They add words and phrases to the word that modify and describe it. Sometimes synonyms and antonyms of a word are supplied by writers, especially when a word in question is intended to convey an abstract or controversial meaning. Writers of prose even resort to the use of poetic or figurative language in their attempts to get over to the reader the precise connotation for a word they have in mind.

These efforts demonstrate the fact that different writers may use a single word to refer to significantly different things or concepts. Knowing this, the child must be alert to infer or make calculated guesses from the context of a passage what particular shade of meaning of a word its author wishes it to have. He can never be satisfied with a static, generalized, universal or surface interpretation of the meaning of a word although such in-terpretations of a word do act as the starting point for the child who then must work to find the specific, localized, and deep structure meaning of a word in question its author wanted to express.

It has even been discovered of late the extent to which the context in which a word is found affects its recognizability by children. Goodman[5] had children read words aloud from word lists, and noted their mistakes. Then he had them read, in sentences, the words they miscalled from this prior reading from lists. He found first graders read correctly in sentences almost two out of three of the words they had misread when reading from lists. Second graders likewise read correctly in sentences three out of four previously missed words. By the third grade, children in this experiment correctly read in sentences four out of five words they earlier had missed when reading from lists. These data strongly suggest that reading by young children is negatively affected when words are out of context. To be suc-cessful readers these children apparently need to decide what peculiar meaning of a word is likely called for, something only the context in which a word is found can explain.

Since both linguists and educational researchers attest that a child's ability to gain the meaning intended for a word depends to a great extent on its semantical variation, we can no longer ignore semantics when making up word lists to be used in the teaching of reading. One way to include semantics in the makeup of basic word lists is to supplement the factor of frequency of occurrence (the basis on which these words are now arranged) with an aspect of semantics that is quantifiable. One matter of semantics that can be counted, and therefore is adaptable to the kind of reckoning

presently done with word lists, is the total number of denotations or separate meanings given words by a large dictionary. It seems purposeful to ask: to what extent would the rankings of words based on their frequency of occurrence be changed if the number of meanings given them by a dictionary was combined with their frequency rankings?

To answer this question I added the number of different meanings given by a large dictionary[6] to each of the ranks of the 200 most frequently occurring words in a recent word list.[7] For example, in this word list *the* was ranked number one. It was found to occur in writings more often than any other word. The large dictionary I used indicated *the* had fourteen different meanings. So, according to my plan *the* was assigned the adjusted score of 15 (1 + 14 = 15). The second most frequently occurring word in the word list I used, *of*, had sixteen dictionary meanings. Its adjusted score therefore was 18 (2 + 16 = 18).

To determine the number of dictionary meanings to assign to a word from the word list I used, the following stipulations were kept in mind: a plural word formed with *s* was assigned the number of meanings given its singular form; a past tense word was assigned the number of meanings given its present tense form, plus any unique meanings it had as a past tense form; comparative and superlative forms were handled in the same way as past tense words; and since many words ending in *ing* were not given a separate heading in the dictionary, they were dropped.

To follow are the first 200 words from the word list used for this study (with the above stipulations in effect) rearranged in an order determined by their adjusted scores (their frequency rankings plus the number of dictionary meanings given them). In parenthesis after each word is given its frequency ranking:

1. and (3)	20. or (26)	39. them (52)	57. into (61)
is (7)	21. were (34)	40. can (38)	will (46)
3. was (13)	22. I (24)	41. said (43)	59. other (60)
4. the (1)	with (17)	she (54)	two (65)
5. he (11)	24. an (39)	these (58)	61. no (71)
6. of (2)	in (6)	44. by (27)	62. more (74)
7. a (4)	26. as (16)	45. many (55)	63. could (70)
are (15)	your (40)	46. has (62)	its (76)
9. his (18)	28. their (42)	what (32)	than (73)
10. they (19)	29. we (36)	48. have (25)	66. my (80)
11. you (8)	when (35)	49. all (33)	67. who (77)
12. it (10)	31. but (31)	then (53)	68. been (75)
13. that (9)	for (12)	51. some (56)	so (57)
14. from (23)	33. each (47)	52. her (64)	70. would (59)
this (22)	if (44)	how (49)	71. now (78)
16. at (20)	35. there (37)	54. him (67)	72. do (45)
not (30)	36. one (28)	55. had (29)	73. may (89)
18. be (21)	which (41)	about (48)	74. first (74)
to (5)	38. on (14)		75. only (85)

76. like (66)
77. see (68)
78. very (93)
79. people (79)
80. find (87)
81. much (103)
82. after (94)
    just (96)
    our (109)
85. know (99)
    most (98)
    where (97)
88. me (111)
    use (88)
90. little (92)
91. too (113)
92. before (104)
93. out (51)
94. any (114)
    new (107)
96. also (120)
97. same (116)
    through (101)
99. another (122)
00. write (108)
01. day (115)
    water (90)
    way (86)
04. because (129)
    did (83)
    three (126)
    up (50)

108. long (91)
109. down (84)
110. time (69)
     used (110)
112. over (82)
     man (112)
114. why (138)
115. must (128)
116. think (119)
117. different (143)
     such (135)
119. again (145)
     here (136)
     years (142)
122. around (121)
123. look (118)
     word (127)
125. get (100)
     good (106)
127. make (72)
128. called (95)
     even (132)
130. away (144)
     every (154)
     things (139)
133. help (140)
     Mr. (160)
135. back (102)
136. well (134)
137. say (152)
     should (159)
139. part (131)

140. great (150)
141. between (157)
     made (81)
143. tell (151)
144. big (162)
     never (171)
146. old (148)
     us (172)
148. come (124)
149. came (123)
     own (167)
151. does (130)
     work (125)
153. right (117)
     still (156)
155. small (153)
156. found (155)
     place (133)
158. might (177)
     name (159)
     next (178)
     put (141)
     while (176)
163. both (184)
     something (182)
165. air (164)
     along (175)
     home (161)
168. always (186)
     few (185)
170. below (180)
     number (149)

172. often (190)
173. off (146)
     under (168)
175. go (105)
176. read (169)
177. last (170)
178. together (191)
179. large (189)
     left (173)
181. asked (192)
182. until (200)
183. going (196)
     important (199)
185. want (197)
186. end (174)
187. give (163)
     school (198)
189. world (195)
190. saw (181)
     thought (183)
192. looked (187)
193. show (188)
194. sound (179)
195. house (193)
196. went (147)
197. don't (194)
     line (165)
199. take (137)
200. set (166)

This rearranged list of basic words indicates that the input of only one factor or semantics (the number of dictionary meanings) causes changes in the rankings of the words determined from their frequency of occurrence alone. For example, *made* moved from rank 81 on its frequency list to 141 on my adjusted list, *up* changed from rank 50 to 104, *had* from 26 to 55. It becomes clear that the addition of a single aspect of semantics can radically change the rankings given words in basic word lists, as they are now constructed.

The discrepancy shown here between a word's ranking based on its frequency of occurrence alone, as versus this plus the number of meanings given it, can be demonstrated in yet another way. I obtained a rank correlation coefficient[8] between two sets of ranks of the 200 basic words I studied: (a) their frequency of occurrence rankings, and (b) their number of different meanings rankings. The degree of correspondence between these

two variables is very low, as indicated by the positive correlation of only .10, a figure that suggests an "almost negligible relationship" between them.[9] We can safely say that a basic word's rank in a word list as based on the frequency of its occurrence alone gives us virtually no indication of the number of meanings a large dictionary gives it. These word list rankings therefore give little evidence as to the different kinds of semantical situations a child potentially will face in attempting to read a given word.

It becomes apparent, as well, that the adjustment made here to the ordering of words on the basic word list I used is only one way word counts such as this can be treated with semantics. Knowing that some basic words have a larger number of semantic differentials than do others could be of help in deciding what constitutes the readability of a written passage. This is but a preliminary step in the reformation we should give the current lists of basic words. The next stage of application of semantics to today's basic word lists might be to determine which dictionary meanings given these words are the same (or at least highly common) meanings held by young children. As Wardhaugh[10] puts it, "It is far more important to know what speakers share in the semantic realm than what separates them, since any understanding of the latter is completely dependent on the former."

So, beyond a simple calculation of the potential meanings a word could engender in its reader, as I have done here, we need to identify which meanings of which words are commonly known to young children. Or which words have several meanings commonly known to children. Then, we could assign these words greater importance in making up lists of basic words than we would assign frequently appearing words whose various meanings are only partially held in common by beginning readers. It seems logical to assume that a word which appears relatively infrequently but whose various meanings are commonly held by most young children would be less difficult for these pupils to learn to read, all other factors being equal, than would a frequently appearing word which has several meanings that are not known to most children.

That this semantic breadth or depth of words is not adequately tested is in evidence in typical vocabulary tests "requiring only a superficial recognition of its (a word's) closest synonym."[11] Nonetheless, few attempts have been made to find if certain meanings of words are understood better by children than others.[3] Unfortunately, what evidence there is appears somewhat contradictory. Russell and Saadeh,[12] for example, discovered that in grade three children chose "concrete" definitions of words significantly more often than "abstract" definitions of them. By grade six the reverse of this was found to prevail. Lundsteen's findings[13] were different from this. The third graders who read her experimental "choose a meaning test" chose "best" definitions of words equally from among the "abstract," "functional," and "concrete" lists of definitions Lundsteen provided in her test. This test was made up of isolated sentences. On the other hand, she found these children chose significantly more "abstract" than "concrete" definitions for words on her experimental "creative and critical paragraph test." This suggests the meanings for words in

paragraphs given to them by children will be more intellectually complex than the meanings they give to words in isolated sentences.

As Lundsteen[13] rightly says, "the first function of reading instruction is not to impart the maximum number of facts, but to develop a repertory of various kinds of meanings, and to encourage use of them in seeking and finding answers to problems." To what extent does the manner in which current basic word lists are compiled contribute to the accomplishment of this crucial goal? The present discussion bears witness that present basic word lists serve this objective very little, if at all. Accordingly, the semantic components of words must be involved in the construction of future word lists if we are to continue to honor them as "basic" to the first function of instruction in reading. It is now almost forty years since Lorge demonstrated that relative frequency "is but a small part of the information needed about words."[14] We should no longer allow his advice to go unheeded.

REFERENCES

1. Robert L. Hillerich. "Word Lists—Getting It All Together." *Reading Teacher,* 27 (January 1974) 353-360.
2. Patrick Groff. "Long Versus Short Words in Beginning Reading." *Reading World,* in press.
3. Edgar Dale, Taher Razik and Walter Petty. *Bibliography of Vocabulary Studies.* Columbus: Ohio State University, 1973.
4. Irving Lorge. *The Semantic Count of the 570 Commonest English Words.* New York: Columbia University, 1949.
5. Kenneth S. Goodman. "A Linguistic Study of Cues and Miscues in Reading." *Elementary English,* 42 (October 1945) 639-643.
6. Jess Stein (editor). *The Random House Dictionary of the English Language.* New York: Random House, 1967.
7. John B. Carroll, Peter Davies and Barry Richman. *The American Heritage Word Frequency Book.* Boston: Houghton Mifflin, 1971.
8. Sidney Siegal. *Nonparametric Statistics.* New York: McGraw-Hill, 1956, pp. 213-223.
9. J. P. Guilford. *Fundamental Statistics in Psychology and Education.* New York: McGraw-Hill, 1950, p. 165.
10. Ronald Wardhaugh. *Reading: A Linguistic Perspective.* New York: Harcourt, Brace and World, 1969, p. 90.
11. David H. Russell. *The Dimensions of Children's Meaning Vocabularies in Grades Four through Twelve.* Berkeley: University of California, 1954, p. 372.
12. David H. Russell and Ibrahim Q. Saadeh. "Qualitative Levels in Children's Vocabularies." *Journal of Educational Psychology,* 53 (August 1962) 170-174.
13. Sara W. Lundsteen. "Levels of Meaning in Reading." *Reading Teacher,* 28 (December 1974) 268-272.
14. Irving Lorge. "The English Semantic Count." *Teachers College Record,* 39 (October 1937) 65-77.