DOCUMENT RESUME

ED 302 836 CS 009 488

AUTHOR Willinsky, John
TITLE The Vocabulary of Cultural Literacy in a Newspaper of Substance.
SPONS AGENCY Canadian Social Science and Humanities Research Council, Ottawa (Ontario).
PUB DATE Nov 88
GRANT 499-87-0033
NOTE 17p.; Paper presented at the Annual Meeting of the National Reading Conference (38th, Tucson, AZ, November 29-December 3, 1988).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Content Analysis; Mass Media Role; Media Research; Newspapers; Popular Culture; Reading Comprehension; *Vocabulary
IDENTIFIERS *Cultural Literacy; Frequency Analysis; Hirsch (E D); *New York Times

ABSTRACT

To examine the role that items on the list of "What Literate Americans Know" (developed by E. D. Hirsch, Jr.) plays in the nation's literacy, a study conducted an electronic search of "The New York Times" to establish the frequency of occurrence for a sample from the list. A random sample of 424 terms (9% of the total list) was selected. Each term or expression which was searched produced a figure representing the frequency of occurrence in the "Times" over a period of 101 months (June 1, 1980 to October 28, 1988), representing a corpus of 660.5 million words. Four frequency periods--yearly, monthly, weekly, and daily--were devised based on how often, on average, a term might be expected to turn up in the newspaper. The terms were also divided into eight categories: geography; history; idioms/proverbs; literature/arts; math/sciences; politics/economics; psychology/anthropology; and religion. Results indicated that any given day's issue of the "Times" contained approximately 2,700 occurrences of terms from the list, with a few of them (such as "New York") making up a good proportion of this number. Geography, the arts, and politics/economics dominated the frequency levels, while history and proverbs/idiomatic expressions were not high frequency categories. Results suggest that Hirsch has identified a corpus of cultural terms which play a part in the daily commerce of the published language. However, to be culturally literate in this set of terms will neither be sufficient nor necessary for a high level of comprehension in reading the "New York Times." (Three tables of data are included.) (MM)

The Vocabulary of Cultural Literacy in a Newspaper of Substance

John Willinsky

Department of Curriculum and Instruction

The University of Calgary

Calgary, Canada, T2N 1N4

Author's Note

The author wishes to thank Paul Frankson, Bryan Green, Carol Malcolm for their professional assitance with this research, and the Social Sciences and Humanities Research Council of Canada (Grant No. 499-87-0033) for its generous support.

The Vocabulary of Cultural Literacy in a Newspaper of Substance

E. D. Hirsch, Jr. (1987) has captured the nation's educational imagination with his concept of cultural literacy. As he tells the story, over the course of this decade he has repeatedly found encouragement to expand on this idea of a literacy based on a shared cultural store of knowledge; his response has grown from an essay to a book to a dictionary of "what every American needs to know." The argument is that the common cultural store of knowledge, necessary for effective communication in a democracy, is not being taught in the schools which are focusing on empty skills and ignoring the substance of the culture, its history, literature, arts, and sciences.

In support of a renewed cultural curriculum for the schools, he and two of his colleagues developed a 5,000 item list of "What Literate Americans Know" (Hirsch, Kett, & Trefil, 1987). The list is "intended to illustrate the character and range of the knowledge literate Americans share" (p. 146). The terms on it are meant to represent the common language which makes communication possible on a national scale. The list would seem about to become a powerful educational tool. The original list was validated by submitting it to "over one hundred consultants" who arrived at a 90 percent agreement on the inclusion of the terms (p. 146).

In order to gain a more realistic picture of the role which the items on the list plays in the nation's literacy, I have conducted an electronic search of The New York Times to establish the frequency of occurrence for a sample from the list. This study speaks to both the weak and the strong claims within Hirsch's argument: on a general level, this list of cultural terms is construed as the common language of a national publication; more strongly,

Hirsch contends that communication in this society is held to be dependent on a familiarity with these terms, or "a high level of literacy is important in holding together the social fabric of the nation" (Hirsch, 1988, p. xi).

As a secondary issue, I wish to demonstrate the profile of culture, literacy, and language which can be derived from the use of national periodicals in electronic databases. These resources are now being employed by dictionary editors to search the current word hoard of the language for new developments (Willinsky, 1988; Shapiro, 1986). If we are serious about students becoming more familiar with the common language of this culture, the database would seem useful research tool for drawing up a profile of that common and shared language which Hirsch feels is necessary to sustain the culture and the state.

Supporting the need for this investigation has been the publication of the list in an expanded and annotated form in The Dictionary of Cultural Literacy (Hirsch, Kett, & Trefil, 1988). In selecting the terms for the Dictionary, however, Hirsch refined the process by setting three criteria which reflect on the study I conducted: The first was for terms to fall between being too specialized and to general. The second test, and most relevant to this study, was to be able to find the entry in "national periodicals": "We reasoned that if a major daily newspaper refers to an event, person or thing without defining it, we can assume that the majority of readers of that periodical will know what that item is (Hirsch, Kett, & Trefil, 1988, p. ix). For the third measure, they "proposed that cultural literacy is not knowledge of current events, although it can help us understand those events as they occur" (p. ix). The point is that "to become part of cultural literacy, an item must have lasting significance" (p. x).

The choice of the Times to search in this study is highly appropriate because the newspaper plays a crucial role in Hirsch's argument in two senses. Not only is it the storehouse of a common language and knowledge, but it is at the forefront in informing the democratic state. Early in Cultural Literacy, Hirsch cites Thomas Jefferson's famous remark about choosing a newspaper over a government if such a choice had to be made (1987, p. 12). Hirsch claims that the government rests on this newspaper readership and its abilities to make sense of the paper: "All citizens should be able, for instance, to read newspapers of substance" (p. 12). Although reading is proving to be a less prominent activity and the newspaper is showing a declining readership since the second world war (Robinson, 1980), studies are still finding that more time is spent reading the paper that other forms of print (Guthrie and Seifert, 1983; Sharon, 1973/74). It remains a major source for this cultural literacy and the perfect spot to begin testing the claims of a prescribed program in cultural literacy.

Method

The commercial database service NEXIS operated by Mead Data Corporation carries The New York Times in electronic form dating back to June of 1980. These files consist of the entire text of the daily paper, including the book review section and the Sunday magazine, but excluding advertising and classified ads. Although the search is conducted electronically at the speed of light, considerable expense and time is still involved in setting up and conducting the searches. For that reason, a random sample of 424 terms (or 9%) was selected from the original list (Hirsch, Kett, & Trefil, 1987, pp.

152-215). Each term or expression which was searched produced a figure representing the frequency of occurrence in the Times over a period of 101 months, from June 1, 1980 to October 28, 1988, representing a corpus of the English language and the culture of 660.5 million words.

In order to gain a sense of how often a term might be expected to turn up in the Times, four frequency periods were devised based on how often, on average, a term might be expected to turn up in the newspaper. Those terms that had a frequency of between 0 and 100 occurrences can be expected to be found on roughly a yearly basis; those with a between 101 to 436 occurrences, can be expected to turn up on a monthly basis; those that occurred between 437 to 3,058 times are at a weekly rate; and those at or better than 3,059 times might be expected to occur on a daily basis. For example, the term "forty winks" showed up 8 times in the search of 101 months which averages out to approximately once a year or, by these divisions, on a yearly basis.

The Times, as a newspaper of substance, averages approximately 218,000 words a day. Given the research on newspaper reading habits, we can expect adults to spend 35-40 minutes a day on average with the paper (Guthrie and Seifert, 1983; Sharon, 1973/74); a good reader reading the paper an hour a day could cover little more than a tenth of the Times, if not unduly distracted by the advertising copy. Thus, to say that a term of cultural literacy turns up on an average once a yearly could well represents ten years of a devoted reader's experience. The four frequency levels are nothing more than a convenience for giving a human scale to what might otherwise be a mass of difficult ratios, such as 1 in 78.6 million words or 1 in 2,500 hours of reading for a term such a "deism" which had nine occurrences each in the Times, or, on average, yearly use.

To add another dimension to this profile of cultural literacy in the Times, the terms were divided into eight categories to suggest which areas of cultural knowledge might be most profitably attended to in preparing to read a national newspaper. The categories are based on a simplified and collapsed version of Hirsch, Kett and Trefil's 23 divisions which are employed in the Dictionary. Using the Dictionary as a guide, the terms were sorted into the categories of geography, history, idioms/proverbs, literature/arts, math/sciences, politics/economics, psychology/anthropology, and religion.

Results

The frequency of occurrence for tne 424 terms over the course of the 8.4 years of The New York Times ranged from 0, for 13 of the terms, to a high count of 236,2690 for "New York" (with the next highest, "New Jersey," at 40,409). The average number of occurrences for the sample was 1,323. This average applied to the total list of approximately 5,000 terms of cultural literacy suggests that any given day's issue of the Times contains approximately 2,700 occurrences of terms from the list, with a few of them, such as "New York," making up a good proportion of this number. The median number of occurrences was 164 suggesting that a number of very frequently used terms skewed the average, and that the majority of terms could be expected to turn up on no better than a monthly basis if the entire paper were read or less often than on a yearly basis given the average reader's habits.

At the high end of the frequency scale, the top 30 terms had ratings over 4,000 occurrences (Table 1). These included a bevy of fundamental political concepts -- "democracy" (11,707 occurrences), "republic" (13,357),

"parliament" (11,306) -- along with a host of economic and geographical terms, all of them highly linked to current events in what might be termed the geo-econo-political sphere of action. But equally so, religion was well represented with "priest" (8,153), "temple" (8,003), and "the Pope" (6,435), although it should be noted that the words can, on occasion, be used in other contexts. A more specific cultural picture could also be found in a few of the popular terms such as the "Securities and Exchange Commission" (7,821) "CIA" (7,245), and "Madison Avenue" (6,088), as well as "creditor" (5,055) and "entrepreneur" (4,944).

Insert Table 1 about here

At the low end of the frequency scale, there were 30 terms which did not turn up more than twice in the 8.4 years of the Times (Table 2). Many of the no-shows were expressions such as "little strokes fell big oaks" and "give him enough rope and he'll hang himself" which perhaps are too banal for the literary standards on the newspaper (Table 2). Science terms, such as "blackbody radiation," "chordates," and "homeostasis," also did poorly suggesting that they may fall outside Hirsch's criteria for non-specialized terms.

Insert Table 2 about here

In tracing the frequency distribution by category, a number of trends

become apparent (Table 3). Geography held steady as might be expected; it plays a part in almost every article if only as a dateline, with the geocentrism of the Times confirmed. Among the geographical centers, after New York and New Jersey, are London (38,930) and Houston (21,017), followed surprisingly by Peru (11,843), and then Rome (9,888), Asia (8,368), and Vermont (8,289). While it would be impossible to predict individual items, it is interesting to think what in the news over the last decade has brought them to such a standing.

Insert Table 3 about here

History, on the other hand, is not a high frequency category. The important American concept of "frontier" (5,092) and "imperialism (1,107) did play a prominent part, but among historical figures in this sample, Tito (1,055) rank first followed at some distance by Alexander Hamilton (491). Among the terms that were treated as historical by the Dictionary of Cultural Literacy were "Cherokee" (584) and "Shawnee" (212).

Proverbs and idiomatic expressions also failed to play a substantial role in the newspaper. It should be noted that the expressions were searched in part ("When in Rome") as well as a whole ("When in Rome do as the Romans do") to catch allusions to them, although some would still slip through, such as "When in Calgary . . ." But it also follows in this type of search that the occurrence of a string of words will be less likely that a single term.

The arts, including a strong representation from literature, was the most prominent category for the entire set overall, although it fell away in terms

of its representation at only the most frequent level. The two most frequently cited terms were "genre" (3,852) and "motif" (3,160) which is an interesting comment on the critical interests of the newspaper, although one might also point to the popularity of "baritone" (2,641) and string quartet (2,276) to suggest a strong musical interest.

The maths and sciences had a fairly even distribution with a rapid falling away at the high level. The most frequent science terms in the sample was "cell" (7,825), followed by "friction" (2,381), with its frequent metaphorical uses in world affairs, and "Einstein" (2,108) as the most popular figure in science.

While the previous five categories have held their own or slipped in the high frequency levels, politics and economics present the most dramatic clustering at the high end, dominating the daily frequency column; this category comes closest to representing a current events approach to cultural literacy which Hirsch is cautious about, but it also included such timeless terms as "bureaucrat" (2,623) and "creditor" (5,055), as well as the great acronyms "SEC" (7,821), "CIA" (7,245) and "GNP" (4,894).

The two smaller contributors to the list were religion and an amalgam of psychology and anthropology. Although with such small numbers it can be misleading to make too much of the way that religion held its own across the frequency levels, yet as I noted it did have representation among the top 30 terms, while psychology, sociology and anthropology got its boost from the diverse concepts of "tenure" (5,078) and "ghetto" (1,882) as well as the political associations of such terms as "charisma" (818) and "sexism" (607).

Discussion

This profile of the different categories of cultural literacy a sense of scale, and some sense of what it is possible to ascertain about the nature of a cultural literacy represented by a leading newspaper. The results suggest that Hirsch and company have identified a corpus of cultural terms which play a part in the daily commerce of the published language. It is hardly surprising that a newspaper would give greater prominence to politics and economics than to literature and the arts, just as it is to be expected that the reverse is the case for the Hirsch, Kett, and Trefil list. Yet this profile by category and frequency does point out that Hirsch's program includes an element of compensation. The program attempts to compensate for a decline that is not the result of poor education or popular culture. That is, the poor showing of idioms and proverbs, for example, and history less so, suggests that the cultural literacy program exaggerates the place of what has already been allowed to pass within the bulwark of this culture. This also speaks to the limitations in the program's claim to represent the vocabulary of such national cultural institutions as the newspaper of substance. The profile presented here suggests that there are real limits to that affiliation which can help us judge the balance between the prescritive and compensatory nature of the list and its descriptive function of a living culture.

But if we allow, as I think we must, that the study has to some degree vindicated Hirsch's quest for a representative list, it also offers a more realistic view of what is at stake in revamping the curriculum of the schools to incorporate a cultural literacy perspective. While the Times employs a number of the terms every day, and most of them every year, it would still be

difficult to construe from this profile that the paper is dependent on the list to communicate all the news that's fit to print. This is to say, Hirsch's weak claim that the list plays a prominent part in the newspaper of substance receives a degree of substantiation, while the stronger claim that the fabric of the nation, as well as the newspaper, depends on the contents of this list needs to be set in perspective. To be culturally literate in this set of terms will be neither sufficient nor necessary for a high level of comprehension in reading this newspaper of substance, and to lead the public or students to believe that mastering this list opens the door to the Times or to the world of culture which it represents needs to be qualified.

The results of this study highlight the fact that much of the Times is, as Hirsch's criteria suggests, too specialized, too general, and too ephemeral. Yet that speaks to the nature of a multi-literate culture and a multi-cultural literacy, that is, to the ways in which we live within a series of cultures and forms of discourse. The daily newspaper is full of specialized areas of reading interest, whether in books, movies, or sports; it is rich in passing events that swell up in the paper for only a few weeks or months, the Oliver Norths of the news, before they pass into near oblivion. We are each specialized and ephemeral readers, and without those interests, it is hard to imagine reading the Times or any newspaper at all. But equally so, the Oliver Norths are crucial in understanding the state of the democracy at any given moment.

The electronic database version of prominent periodicals has begun to be explored as a research tool into the nature of literacy. The natural extension of this study is to use the database for an analysis of how these terms of cultural literacy function in the context of the articles in which

they appear. Questions can be raised about the degree to which the term contributes to the meaning of a piece, and the ways in which the text can be said to teach the term. It may be, for example, that the best way to learn the cultural literacy of The New York Times is to read it. A second point to be considered is that the database also points to another manner of looking at cultural literacy, at a literacy which is concerned with how language is used on a national scale. The database version of leading periodicals does lend itself to statistical profiles of terms and categories, to tracking the fate and state of the language in point. Instead of learning terms ascertained by a committee of a hundred scholars, students could use such tools to become their own investigators into the current life of the language and the culture.

To begin to reshape the nature of instruction in the schools around a concept of cultural literacy requires that two claims be substantiated. The first is that a defined lexicon of cultural literacy is indeed the lifeblood of effective communication and the backbone of democracy in the nation. The second is that an education in such a form of literacy will enhance the learner's opportunities and participation in this national enterprise. Searching electronic databases of a prominent periodical is a small step, admittedly, in ascertaining the validity of Hirsch's claims and providing a more accurate picture of the literacy that sustains the commonwealth. But there is little question that these claims must be taken seriously and investigated thoroughly by all the means at our disposal, if we are to make wise choices for the future of the school and society.

References

Hirsch, E. D. Jr., (1988). The theory behind the dictionary: Cultural literacy and Education. In E. D. Hirsch Jr., J. F. Kett & J. Trefil. The dictionary of cultural literacy: What every American needs to know (pp. xi-xv). Boston: Houghton Mifflin.

Hirsch, E. D. Jr. (1987). Cultural literacy: What every American needs to know. Boston: Houghton Mifflin.

Hirsch, E. D. Jr., Kett, J. F., & Trefil, J. (1987). What literate Americans know: A preliminary list. In E. D. Hirsch Jr., Cultural literacy: What every American needs to know (pp. 146-215). Boston: Houghton Mifflin.

Hirsch, E. D. Jr., Kett, J. F., & Trefil, J. (1988). The dictionary of cultural literacy: What every American needs to know. Boston. Houghton Mifflin.

Guthrie, J. T., & Seifert, M. (1983). Profiles of reading activity in a community. Journal of Reading, 26, 498-509.

Robinson, J. P. (1980). The changing reading habits of the American public. Journal of Communication, 30, 141-152.

Shapiro, F. R. (1986). Yuppies, yumpies, and computer-assisted lexicography. American Speech, 61, 139-147.

Sharon. A. T. (1973/74). What do adults read? Reading Research Quarterly, 9, 149-169.

Willinsky, J. (1988). Cutting English on the bias: Five lexicographers in search of the new. American Speech, 63, 44-66.

Table 1
The 30 Most Frequently Used Cultural Literacy Terms in The New York Times

Asia	New Jersey
Atlantic City	New York
cell	parliament
Central Intelligence Agency (CIA)	pension
commodity	Peru
Court of Appeals	the Pope
creditor	priest
democracy	republic
dividend	Rome
entrepeneur	Securities and Exchange Commission (SEC)
gross national product (GNP)	Securities Council
Iraq	Switzerland
London	temple
Madison Avenue	Vermont
merger	Warsaw

Note. These 30 terms ($N = 424$) had greater than 4,800 occurrences in *The New York Times* between June 1, 1980 and October 28, 1988.

Table 2
The 30 Least Frequently Used Cultural Literacy Terms in The New York Times

ancien regime	judicial review
Birds of Venus	land breeze
blackbody radiation	law of contridiction
Burr-Hamilton duel	little strokes fell great oaks
chordates	Nero fiddled while Rome burned
an elephant never forgets	nitrogenous wastes
export quote	not with a bang but a whimper
for want of a nail the kingdom was lost	man shall not live by bread alone
homeostasis	moon made of green cheese
I wandered lonely as a cloud	patience of Penelope
indefinite article	poets are born, not made
Indian file	a stitch in time saves nine
Jamestown settlement	Tudor monarchy
John XXIII	walk under a ladder
Joshua Fit the Battle	You cannot serve God and mammon

Note. These 30 terms ($N = 424$) had no more than 2 occurrences in *The New York Times* between June 1, 1980 and October 28, 1988.

Table 3

Percentage Occurrence of Cultural Literacy Terms by Category at Four Frequency Levels

Category	Frequency Level				
	Yearly % (*f*)	Monthly % (*f*)	Weekly % (*f*)	Daily % (*f*)	Total % (*f*)
Geography	3.5 (15)	2.8 (12)	4.5 (19)	4.2 (18)	15.1 (64)
History	6.1 (26)	4.2 (18)	0.9 (4)	0.2 (1)	11.6 (49)
Idioms/Proverbs	9.4 (40)	5.4 (23)	0.7 (3)	0.0 (0)	14.2 (66)
Literature/Arts	7.1 (30)	8.4 (36)	4.5 (19)	0.5 (2)	20.5 (87)
Math/Sciences	7.5 (32)	5.0 (21)	3.8 (16)	0.5 (2)	16.7 (71)
Politics/Economics	0.7 (3)	4.0 (17)	5.0 (21)	5.2 (22)	14.9 (63)
Psych/Anthropology	1.2 (5)	0.5 (2)	0.7 (3)	0.2 (1)	2.6 (11)
Religion	1.2 (5)	0.9 (4)	0.0 (0)	0.9 (4)	3.1 (13)
Total	36.7 (156)	31.4 (133)	20.0 (85)	11.8 (50)	100.0 (424)

Note. Percentages are based on total number of occurrences of all terms ($N = 424$) found in *The New York Times* between June 1, 1980 and October 28, 1988. Frequency levels are based on averaging the number of occurrences of the term over the 101 months of the *Times*: yearly = 0 - 100 occurrences; monthly = 101 - 436 occurrences; weekly = 437 - 3058 occurrences; daily = 3059 + occurrences.