A study extended word association methodology beyond isolated word stimuli to investigate the effects of written context on the meanings that proficient readers impart to words. A repeated-measures design was used to assess the responses of 62 sixth grade readers to target words at three levels: no context, limited context, and expanded context. Two means used to assess the meaning of target words were number of response categories and proportion of paradigmatic responding. The results indicated that meanings attached to a word were more constrained when the word was embedded in a sentence than when it was isolated. The findings refute the implications of earlier word association research. (Author/FL)
Word Association Extended: A Study of the Effects of Written Context on Word Meaning

Adrienne S. Escoe
Southwest Regional Laboratory for Educational Research and Development
Los Alamitos, California 90720

American Educational Research Association Meeting
Los Angeles: April, 1981
The word association technique has a long history of use in the study of human development and behavior. Most of the efforts have been addressed to the developmental nature of associations between words (Bartel, Grill, & Bartel, 1973; Brown & Berko, 1960; Entwisle, Forsyth, & Muuss, 1964; Ervin, 1963; Francis, 1972; Lippman, 1971; McNeill, 1966; Palermo, 1971).

Some researchers have extended word association studies beyond the investigation of developmental patterns and have applied the technique as a tool for examining word meaning (Deese, 1965; Dinnan, Neilsen, & Crable, 1976; Ervin, 1961; Noble, 1952). The greatest weakness of these efforts, however, has been the failure to recognize the role of context in the impartation of meaning to words. Investigators have, for the most part, neglected to consider that word association research could contribute to an explanation of how word meaning is processed in contextual language, especially in reading. Instead they have tended to limit their work to the study of isolated words. The disregard of context in research methodology becomes critical when researchers attempt to apply their findings to an explanation of reading behavior. These investigations seem to suggest that responding to words in isolation is comparable to responding to words that are surrounded by others, such as would be the case in typical reading situations (Bigaj, Dinnan, & Crable, 1977; Dinnan, Bickley, & Cowart, 1971; Mickelson, 1972; Tobiesyen, 1964). It has been suggested that proficient readers are aware of the specific textual setting of a word and utilize the constraints of context, but that the "constraints" that are actually examined in association studies may bear little relationship to natural language (Weaver, Kingston, & Dinnan, 1971). Is it mainly a matter of historical accident that
association studies have yielded relatively little information toward an understanding of how connected discourse is processed? Furthermore, should researchers be asking whether such information can ever come from the data these studies generate (Fillenbaum & Jones, 1965)?

One of the purposes of this study was to meet a challenge raised by researchers such as Fillenbaum and Jones (1965): to expand the use of the word association technique as a tool for understanding how a reader imparts meaning to words encountered in written language. The technique of responding to word stimuli by free association (i.e., word association) is particularly appropriate for studying word meaning in reading because word association seems to figure prominently in a concept of word meaning as well as in a theory of reading behavior. As part of the study, a concept of meaning was developed that recognized the semantic attributes or features of a word (including syntactic attributes) and the constraints imposed by the surrounding context. There are features of a word that determine its placement in a semantic hierarchy, and these features form a network that represents word meaning (Anderson, 1972; Quillian, 1968). That is, how different words are related in terms of meaning is a function of the connections among the attributes of the words themselves. Further, assessments of word relatedness have been provided by the use of word association methodology in classic studies of word meaning. The technique has been applied, however, only to words in isolation (Deese, 1962; Noble, 1952). Other research has focused on the associational aspect of context in the impartation of meaning to words. Hörmann (1971), for one, suggested that context momentarily strengthens certain associations of a word, which has the effect of making other meanings less available to
the reader. An earlier study by Hové and Osgood (1954) suggests that
word-association is a function of the probability that a certain word will
appear in various contexts.

The data from word-association research can support not only a concept
of word meaning, but a theory of reading behavior as well. There is a
growing consensus among theorists that reading behavior is an hypothesis-
testing process, or a cognitive search for meaning and its verification.
Within this theory it can be inferred that the associations of a word by
means of its attributes and the constraints of the context surrounding
the word contribute to readers' information-seeking and verifying pro-
cesses by which hypotheses of meaningfulness for words encountered in
reading are tested.

Supported by a concept of meaning that recognizes word attributes
and contextual constraint and by an hypothesis-testing theory of reading,
word association methodology was used to generate two hypotheses in this
study. One hypothesis was based on the assumption that responses to
target words could be assigned categories of meaning that would reflect
various sets of attributes for the word. For example, given the target
word 'bed', three responses were a 'flowerbed', a 'place to sleep', and something
to sleep on. The features of the first response would suggest assignment
to one category. The features of the second and third responses would
suggest another category of meaning. The number of different categories
of response to a target word, or range, was considered to be an indication
of the meaning imparted to the word. Features that are supplied by
context were seen as constraining the range of word meaning. It was
hypothesized that the number of response categories would decrease as the
of the written context surrounding the word increased. A second hypothesis was based on the notion that a paradigmatic response to a word, that is, same form class as target word (e.g., *king* - *libn*), was more likely to share semantic features with the word than a non-paradigmatic response (e.g., *king* - *mad*). Paradigmatic responding to a word was considered to be another indication of the meaning imparted to a word. It was hypothesized that the proportion of paradigmatic responses to a word would increase as the size of the written context surrounding the word increased.

**Method**

**Subjects**

Sixty-two proficient sixth-grade readers participated in the study. Subjects were drawn from the entire sixth grade of one elementary school (about 100 students) and from one class of another school (about 20 students). Both schools were located in a suburban county of the state of Maryland. Students were not included in the study if they did not score at least 5.0 on the reading comprehension subtest of the Iowa Tests of Basic Skills administered one year prior to data collection or did not receive a positive rating on a scale developed by the investigator to provide a judgment of a reader's comprehension of a variety of written materials over several months' time. Each student included as a subject met additional criteria: did not repeat a grade; responded satisfactorily to all instruments, judged according to specified criteria (see Escoe, 1979); and received written parental permission to participate in the investigation.
**Materials**

The following instruments were developed by the investigator: an association test of isolated words, an association test of words in limited context, and an association test of words in expanded context. The test of isolated words consisted of 21 nouns, including 10 target words that would be used to analyze effects of context. The test of limited context contained 16 single sentences, 10 of which included the same target words as the test of isolated words. The test of expanded context was composed of 11 sets of three sentences. Across sets the second sentences consisted of the 10 target sentences of the test of limited context plus one of the remaining six filler sentences drawn at random.

The 10 target words were selected by the following procedure to maximize equivalence of the target words. First, all nouns appearing on the Palermo and Jenkins (1964) list of 200 stimulus words were identified. Only nouns were used so that the findings of this study could be compared with prior research such as that of Mickelson (1972) and others which had drawn items from the same source. If a noun was classified as more than one part of speech, it was drawn for the word sample only if it was classified first as a noun according to The Random House Dictionary of the English Language (Stein, 1969), which lists the most frequent classification first. Next, the resultant pool of 74 words was checked by two lists of frequency data to ensure that the words occurred in fifth- and sixth-grade reading material (Carroll, Davies, & Richman, 1971; Thorndike & Lorge, 1944). The remaining 66 words were rated by six adults on a scale developed by the investigator to assess the imagery value of the words. The instrument was designed according to the theory that concrete
nouns are superior to abstract nouns in their capacity to elicit imagery, and that imagery can facilitate the formation of association between words (Paivio, 1965). Words rated as easy to picture were selected, because they would be more likely to facilitate responding. Finally, the group of 54 highly picturable words was narrowed to those words which received the 10 highest number of dictionary entries listed for the word, using *The Random House Dictionary*. Target words that had the most entries were seen as supplying the greatest range of meaning categories.

The test of isolated words contained 11 filler words in addition to the target words. The filler words were chosen randomly from all words that met the criteria described above for target words, except for imagery value and number of dictionary entries (i.e., they were selected from 56 words that remained after target words were removed from the list of 66 words). Filler words served training and procedural purposes. The tests of limited context and expanded context contained filler sentences for the same reasons. Each sentence generated for the limited context met additional criteria: was embedded with a stimulus word (target or filler) that was used as a noun; was generated with comparable syntax in terms of both structure and complexity according to Endicott's scale (1973); was generated according to "objective" case relations (Fillmore, 1968); was comprised entirely of words that met the Thorndike-Lorge (1944) and American Heritage (Carroll, et al., 1971) criteria specified for the selection of target words; and received a positive rating of acceptability from each of five graduate students in reading. Acceptability was based on whether the sentence was likely to be included in children's reading materials; that is, did not sound awkward or contrived. Each set of
sentences generated for the expanded context was constructed according to the general pattern for the limited context with additions: a sentence from the test of limited context was the second sentence within a set, and the first and third sentences in a set reflected "agentive" case relations (Fillmore, 1968).

Each instrument was a stapled booklet printed with one item per page. Filler items were used to equalize responding time among the three instruments. The order of all items, including targets, was varied randomly for each booklet; however, the positions of target and filler items were fixed. For training purposes, a filler item appeared on the first page of each booklet. To maximize on-task reading behavior, all items for the tests of limited context and expanded context were preceded by a restricted cloze task, an omitted word for each sentence determined randomly, applied to the same item but on a separate page. Sample target items are presented in Table 1.

Insert Table 1 about here

Tasks.

Subjects were asked to respond to two tasks for each instrument: Task 1 consisted of varied-length responses (single word, phrase, sentence) to approximately half of the items, including five target words. Task 2 was restricted to single-word responses to the other test items, including the other five target words. The target words were:

<table>
<thead>
<tr>
<th>Task 1</th>
<th>Task 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>man</td>
<td>hand</td>
</tr>
<tr>
<td>bed</td>
<td>king</td>
</tr>
<tr>
<td>numbers</td>
<td>foot</td>
</tr>
<tr>
<td>square</td>
<td>light</td>
</tr>
<tr>
<td>head</td>
<td>house</td>
</tr>
</tbody>
</table>
The filler words were: *lamp* (for each of the three tests); *boy*, *bible*, *stove*, *city*, and *tobacco* (for the tests of isolated words and limited context); and *chair*, *river*, *anger*, *carpet*, and *religion* (for the test of isolated words only). Task 1 responses were analyzed in terms of the number of different meaning categories that were assigned to the responses by raters, while Task 2 responses were analyzed in terms of raters’ designations of paradigmatic or non-paradigmatic. Estimates of interrater reliability were established among the three raters for each task. For Task 1 responses, both percentages of agreement and Kappa coefficients of agreement between pairs of raters were provided. Reliability estimated between each pair of raters, using percentages of agreement, ranged from 61.83 to 98.92 for individual target words, with a mean across all words and all pairs of raters at 88.21. With Kappa coefficients the range was from .445 to .981 with a mean of .827. For Task 2 responses, the Phi coefficient was used: interrater reliability was estimated at .798 to 1.000 for individual target words, with a mean of .954 across all words and all pairs of raters.

Procedure

Tests were administered individually to each subject on each of three days spaced one week apart. The administration of the instruments was ordered in six different sequences. Subjects were randomly assigned a sequence of test administration. For Task 1 items, subjects were asked to tell what the target word made them think of; for Task 2, they were asked to tell in one word what the target word made them think of. Responses were recorded in writing by the test administrator.
Design

A repeated-measures design was used for a single group of subjects. The effect of context on word meaning was assessed in two ways. One was through the number of different categories of meaning assigned by a rater as corresponding to responses to a word. A category number was assigned to each Task 1 response elicited under the three conditions of context, each response listed in order of data collection. No indication was given as to which association test was administered when the particular response was given. Category numbers were assigned from lists of dictionary entries selected according to specified criteria (see footnote 2) and tallied for each target word under each condition of context. The index of dispersion (D) was employed to assess the variation of response categories among the three conditions of context.

Another assessment of the effect of context on word meaning was provided by proportions of paradigmatic to total responses. A designation of paradigmatic or non-paradigmatic was assigned by a rater for each Task 2 response, listed in order of data collection. The particular test administered for any response was not indicated. The number of paradigmatic responses to each word was tallied and averaged and transformed into a mean proportion of total responses.

Results

Index of dispersion (D) values representing the number of different categories corresponding to subjects' responses to the five target words under three levels of written context are presented in Table 2. Examination of the table indicates that the number of different response categories
decreased from the condition of a context of no words (isolated words) through that of three sentences (expanded context). Analysis of variance procedures (ANOVA) for repeated measures resulted in a decision to reject an hypothesis of no differences based upon the conservative F test (Geisser & Greenhouse, 1958); $F(1,4) = 26.38, p < .01$, $MSe = .022$. The decision also held with the conventional F test (Dayton, 1970; Romaniuk, Levin, & Hubert, 1977), $F(2,8) = 26.38, p < .01$, $MSe = .022$. Using the Newman-Keuls post hoc procedure with all possible pairwise comparisons at the .01 level, significant differences were found between the conditions of isolated words and single-sentence context and between isolated words and three-sentence contexts. The difference between single-sentence and three-sentence context, however, was found to be not significant.

The mean proportions of paradigmatic to total responses to five target words under three conditions of written context are presented in Table 3. It may be seen from Table 3 that the mean proportion of paradigmatic to total responses for each target word and for the total target words increased from the condition of isolated words through the condition of a three-sentence context. Unlike the data collected to test the first hypothesis, these data provided frequency scores; therefore, scores of the mean proportions, and not D values, were the units employed in analysis of variance procedures. ANOVA for repeated measures resulted in significant differences under the three conditions of context for the main effect with
both conservative and conventional $F$ tests, $F(1, 61) = 377.70, p < .01$, MSe = .084 and $F(2, 122) = 377.70, p < .01$, MSe = .084, respectively. On the basis of the $F$ tests, however, no decision was reached as to whether the effect of content (words) or the interaction of size of context and content was significant because of conflicting results of analyses, for both effects a conventional, $F$ test resulted in the rejection of an hypothesis of no differences, $F(4, 244) = 5.20, p < .01$, MSe = 180 (content) and $F(8, 488) = 3.91, p < .01$, MSe = 136 (context x content), while the conservative $F$ test resulted in failure to reject an hypothesis of no differences, $F(1, 1) = 5.20, p < .01$, MSe = 180 (content) and $F(1, 61) = 3.91, p < .01$, MSe = 136 (context x content). An Attempt to test for homogeneity of variance: covariace was unsuccessful due to a vanishing determinant and a singular matrix in the computational procedures. The main effect of context and the effect of content were investigated further with all possible pairwise comparisons. Using the Newman Keuls post hoc procedure at the .01 level, all differences of main effect were found to be significant. All comparisons, except ones of content were found to be not significant. Investigating the interaction of size of context and content further with the Scheffe post hoc procedure at the .01 level (using only the largest differences between the differences between content means in the contrasts), no interactive effect was found to be significant.

Though a part of the study was to determine the effect of written context on word meaning, the approach used...
served an additional research purpose, which was to extend the word association technique as a tool for understanding how meaning is imparted to words encountered in reading situations.

The finding that the size of the written context surrounding a word had a significant effect on the number of different response categories corresponding to the word appears to support the notion of reading as an hypothesis-testing behavior as advanced by Goodman (1967), Hall & Ribovitch (1973), Pearson (1978), Samuels (1970), and Smith (1978). It was predicted that the presence of a larger size of context contributes more information to assist the language user in attaching appropriate meaning to a word in reading. However, post hoc analysis indicated that the constraint in choice of meaning provided by a three-sentence context was not significantly greater than that provided by a single sentence one. A plausible interpretation of this finding is that given the limitations of studying only three conditions of context, a typical written sentential context (clearly not an isolated word) prior to and following a word tends to serve the reader in hypothesis testing by providing him or her with enough features that may be associated with the word. Of course, if sentences are constructed accurately, to create an equally as context a to if the case in language research, different results could be expected. The sentential contexts used in the study were intended to reflect with graders' real reading material

The finding that a three sentence three-sentence context by itself, in response, is responses than the size just shorter than it supports the claim expressed by Hörmann (1971), that written text at the provide features to assist
the reader in limiting the meanings he or she attaches to a word. The support is based on the notion that paradigmatic associations are more likely to reflect shared features of words (Francis, 1972). Why one indication of word meaning, that is categories of response, and a second indication, paradigmatic responding, did not yield wholly consistent results is subject to further study. One explanation, albeit a premature one, is that single-word responses (Task 2) to determine paradigmatic/non-paradigmatic associations to word stimuli produced a more restricted indication of meaning than multiple word responses (Task 1).

From the data gathered here it seems reasonable to conclude that the many researchers who used the word association technique with isolated words may have overgeneralized when they applied their findings to implications concerning contextual-language behavior, especially reading. For example, Mickelson (1972) studied the relationship between associative verbal encoding, a measure of fluency in responding to isolated words, and reading achievement. She found that improved associative verbal encoding resulted in better reading performance. From the results of her study, Mickelson concluded that fluency in responding to words could seem to be an essential component in an hypothesis testing theory of reading comprehension. While a readily available repertoire of isolated word may be an indication of an individual's competence in any linguistic process (including reading), a notion that Mickelson describes to nonlinguists, it is difficult to accept this notion as support for the of reading comprehension in the absence of appropriate investigation of language units longer than a single word. Acknowledging the limitations of the present study, the results attest that responses to isolated word stimuli...
are not the same as responses to words that are embedded in a linguistic context. Further research is suggested to strengthen the conclusions of this study and to widen their applicability: double-blind replication with other populations, larger contexts, other target words; study of the contributions of extra-linguistic context and interaction with linguistic context in the impartation of meaning to words; investigation of instructional applications, such as the effect of context on teaching/learning word meaning.

In the past, word association research has contributed much toward explanations of human development and behavior. It has helped in the study of some language processes. What word association research has failed to do up to now, however, is provide information about word meaning in written contexts. From these data, it appears that the word association technique extended is not only appropriate as a tool for investigating word meaning in reading situations, but also can be a reliable and efficacious means to a crossroads of inquiry, namely meaning and reading.
References


Palermo, D. S. Characteristics of word association responses obtained from children in grades one through four. *Developmental Psychology*, 1971, 5, 118-123.


Footnotes

1Detailed support for these notions and how they contribute to a theory of reading is provided in the author's doctoral dissertation completed at the University of Maryland, 1979.

2For consistency, criteria were established for including/excluding individual entries in the count. Included were: expressions beginning with the stimulus word; proper nouns, titles (e.g., Man and Superman); each numbered entry. Excluded were: affixed words, if the unaffixed word is counted (e.g., heads); possessives; non-English expressions (e.g., man spricht deutsch); lettered items within entries; abbreviations; hyphenated words.
Table 1
Examples of Target Items

<table>
<thead>
<tr>
<th>Test of Isolated Words</th>
<th>Target-Response Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test of Limited Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restricted Cloze Presentation</td>
</tr>
<tr>
<td>Both numbers came, the animal acts</td>
</tr>
<tr>
<td>after statue regular</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test of Expanded Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restricted Cloze Presentation</td>
</tr>
<tr>
<td>Bob tore the music sitting on the stand above the piano. Both numbers came after the animal acts. He walked out the stage door and sang in the empty street.</td>
</tr>
</tbody>
</table>
Table 2

Index of Dispersion (D) Values for Number of Different Response Categories to Task 1 Words by Context

<table>
<thead>
<tr>
<th>Target Word</th>
<th>Context</th>
<th>Isolated Word</th>
<th>Limited</th>
<th>Expanded</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>man</td>
<td></td>
<td>.878</td>
<td>.709</td>
<td>.184</td>
<td>.591</td>
</tr>
<tr>
<td>bed</td>
<td></td>
<td>.864</td>
<td>.648</td>
<td>.313</td>
<td>.608</td>
</tr>
<tr>
<td>numbers</td>
<td></td>
<td>.852</td>
<td>.452</td>
<td>.418</td>
<td>.574</td>
</tr>
<tr>
<td>square</td>
<td></td>
<td>.879</td>
<td>.292</td>
<td>.032</td>
<td>.401</td>
</tr>
<tr>
<td>head</td>
<td></td>
<td>.916</td>
<td>.125</td>
<td>.032</td>
<td>.566</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>.877</td>
<td>.451</td>
<td>.196</td>
<td>.508</td>
</tr>
</tbody>
</table>
### Table 3
Mean Proportions of Paradigmatic to Total Responses Across Task 2 Words by Context

<table>
<thead>
<tr>
<th>Target Word</th>
<th>Context</th>
<th>Isolated Word</th>
<th>Limited</th>
<th>Expanded</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>hand</td>
<td></td>
<td>.419</td>
<td>.694</td>
<td>1.000</td>
<td>.704</td>
</tr>
<tr>
<td>king</td>
<td></td>
<td>.210</td>
<td>.114</td>
<td>.984</td>
<td>.656</td>
</tr>
<tr>
<td>foot</td>
<td></td>
<td>.307</td>
<td>.114</td>
<td>1.000</td>
<td>.654</td>
</tr>
<tr>
<td>light</td>
<td></td>
<td>.258</td>
<td>.548</td>
<td>.984</td>
<td>.591</td>
</tr>
<tr>
<td>house</td>
<td></td>
<td>.584</td>
<td>.120</td>
<td>1.000</td>
<td>.790</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>.355</td>
<td>.110</td>
<td>.994</td>
<td>.688</td>
</tr>
</tbody>
</table>