Use this as the main Sternberg rdg!

THE NATURE
of
VOCABULARY ACQUISITION

Edited by

Margaret G. McKeown • *Mary E. Curtis*

University of Pittsburgh

Harvard University

6 Most Vocabulary is Learned From Context

Robert J. Sternberg
Yale University

The thesis of this chapter is simple: Most vocabulary is learned from context. This chapter documents this claim, and draws out some of its main implications. Of equal importance, I discuss what the claim does not imply. What the claim does imply is that teaching people to learn better from context can be a highly effective way of enhancing vocabulary development. What the claim does not imply is that teaching specific vocabulary using context is the most effective, or even a relatively effective, way of teaching that vocabulary. Unfortunately, many believers in learning from context, as well as their detractors, have drawn the second inference rather than the first. As a result, they are on the verge of throwing out a perfectly clean and healthy baby with its, admittedly, less than sparkling bath water.

My chapter is divided into five basic parts. First, I present what I consider to be three basic facts about vocabulary. Second, I attempt to account for these facts within a single explanatory framework. Third, I discuss the implications of the facts and framework for the teaching and learning of vocabulary. Fourth, I provide an overview of some data we have obtained using these facts and framework as a basis for teaching vocabulary-acquisition skills. Finally, I draw some conclusions.

THREE BASIC FACTS ABOUT VOCABULARY

Let's start with the three basic facts: First, although estimates of adult vocabulary sizes differ, there is no disagreement that the typical adult has in his or her vocabulary tens of thousands of words, and in exceptional cases,

adults may have vocabularies in excess of 100,000 words. Second, one's level of vocabulary is highly predictive, if not determinative, of one's level of reading comprehension. Despite the gross surface dissimilarity between vocabulary tasks, which seem primarily to measure acquired knowledge, and reading comprehension tasks, which seem primarily to measure the ability to acquire new knowledge, the correlations between these two types of tests often are almost as high as their reliabilities. Third, vocabulary is probably the best single indicator of a person's overall level of intelligence. Stated in another way, if one wants a quick and not-too-dirty measure of a person's psychometrically measured intelligence, and thus has time to give just one brief test of it, vocabulary is generally the best predictor of overall score on a psychometric IQ test.

These are three basic facts about vocabulary. It would be theoretically elegant, at best, and parsimonious, at worst, if we could account for all three of these facts in terms of a single explanatory mechanism, or theory.

AN EXPLANATORY FRAMEWORK
FOR THE THREE BASIC FACTS

There is a simple explanation to account for all three of these facts. The explanation is that most vocabulary is learned from context. During the course of one's lifespan, one is exposed to innumerable words through seemingly countless sources—textbooks, lectures, newspapers, magazines, friends, enemies, parents, movies, and so on. Even if one learned only a small proportion of the words thus encountered in the contexts in which they are presented, one could plausibly develop a vocabulary of tens of thousands of words, which represents only an infinitesimal proportion of our exposure to words. It is difficult to see any other way in which so many words could be learned. Certainly, no one has sat down and memorized tens of thousands of words, or formed keyword images linking new words, old words, and pictures. Nor are memorization or forming of keywords practical strategies for learning a language as a whole, as opposed to selected words within that language. There is just too much to learn! Context can also help explain the link to reading comprehension: Whereas vocabulary is an indirect measure of ability to learn word meanings in context, reading comprehension is a fairly direct measure of ability to learn concepts in context, where the concepts may be single words, facts, ideas, or whatever. The major difference would then be that reading comprehension tests measure present ability to learn from context, whereas vocabulary tests measure past ability. Except in the cases of old age or mental infirmities, where present ability may be significantly less than past ability, one would expect these two measurements to be highly correlated. The contextual view also explains the high correlations with intelli-

gence test scores. Much of intelligence is, quite simply (as well as complexly!), the ability to learn from context. A better learner of anything, not just of words, is someone who can use context to increase his or her knowledge base, which will later form the basis for expertise in any of a variety of content areas.

Just how does one learn vocabulary from context? Sternberg and Powell (1983) have proposed that learning from context involves three basic ingredients: knowledge-acquisition processes, contextual cues, and moderating variables. Consider, for example, the passage fragment:

> Although for the others the party was a splendid success, the couple there on the blind date was not enjoying the festivities in the least. An *acapnotic*, he disliked her smoking; and when he removed his hat, she, who preferred "ageless" men, eyed his increasing *phalacrosis* and grimaced.

Processes of Knowledge Acquisition

According to our theory (Sternberg & Powell, 1983), three processes are applied to learning new words such as *acapnotic* and *phalacrosis*.

1. *Selective encoding* involves separating relevant from irrelevant information for the purposes of formulating a definition. For example, it is relevant to figuring out the meaning of acapnotic that he disliked her smoking, but it is not relevant that the couple was at a party.

2. *Selective combination* involves combining relevant cues into a workable definition. For example, in figuring out the meaning of phalacrosis, one needs to combine at least four relevant cues—that the woman likes "ageless" men, that the removal of the man's hat led to the conclusion that the man is not ageless, that the man with whom she has the date is showing increasing signs of phalacrosis, and that she grimaced as a result.

3. *Selective comparison* is a process by which new information about a word is related to old information already stored in memory. To figure out the meaning of phalacrosis, for example, one needs to know that among men, baldness often increases with age, and even that baldness becomes visible when one removes one's hat.

Contextual Cues

The three processes of selective encoding, selective combination, and selective comparison do not operate in a vacuum, or at random. Rather, they operate on a relatively stable set of cues provided by the context in which new words occur. These cues include:

1. *Temporal cues*: cues regarding the duration or frequency of X (the unknown word), or when X can occur

2. *Spatial cues*: cues regarding the location of X, or possible locations in which X can sometimes be found

3. *Value cues*: cues regarding the worth or desirability of X, or the kinds of affects X arouses

4. *Stative descriptive cues*: cues regarding properties of X (such as size, shape, color, odor, feel, etc.)

5. *Functional descriptive cues*: cues regarding possible purposes of X, actions X can perform, or potential uses of X

6. *Causal/Enablement cues*: cues regarding possible causes of or enabling conditions for X

7. *Class membership cues*: cues regarding one or more classes to which X belongs, or other members of one or more classes of which X is a member

8. *Equivalence cues*: cues regarding the meaning of X, or contrasts (such as antonymy) to the meaning of X.

Moderating Variables

Because of moderating variables, it is not always equally easy to apply the three knowledge-acquisition processes to the eight context cues. These variables make it either easier or harder to apply the processes to the cues. The moderators include variables like the following:

1. *Number of occurrences of the unknown word*: multiple occurrences of an unknown word increase the number of available cues and can increase the usefulness of individual cues if readers integrate information obtained from cues surrounding the multiple occurrences of the word. For example, the meaning of a given temporal cue may be enhanced by a spatial cue associated with a subsequent appearance of the unknown word, or the temporal cue may gain in usefulness if it appears more than once in conjunction with the unknown word. On the other hand, multiple occurrences of an unfamiliar word can also be detrimental if the reader has difficulty integrating the information gained from cues surrounding separate appearances of the word, or if only peripheral features of the word are reinforced and are therefore incorrectly interpreted as being of central importance to the meaning of the unfamiliar word.

2. *Variability of contexts in which multiple occurrences of the unknown word appear*: different types of contexts, for example, different kinds of subject matter or different writing styles, and even just different contexts of a given type, such as two different illustrations within a given text of how a word can be used, are likely to supply different types of information about the unknown word. Variability of contexts, up to a certain point, increases

the likelihood that a wide range of types of cues will be supplied about a given word, and thus increases the probability that a reader will get a full picture of the scope of a given word's meaning. Although variability of contexts can help, too much variability can overwhelm the reader and interfere with learning the meaning of a new word. If the information is presented in a way that makes it difficult to integrate across appearances of the word, or if a given individual has difficulties in making such integrations, then the variable repetitions may actually obfuscate rather than clarify the word's meaning. In some situations and for some individuals, a stimulus overload may occur, resulting in reduced rather than increased understanding. Mere repetition of a given unknown word in essentially the same context as that in which it previously appeared is unlikely to be as helpful as a variable-context repetition, because few or no really new cues are provided regarding the word's meaning.

3. *Importance of the unknown word to understanding the context in which it is embedded*: if a given unknown word is judged to be necessary for understanding the surrounding material in which it is embedded, the reader's incentive for figuring out the word's meaning is increased. If the word is judged to be unimportant to understanding what one is reading (or hearing), one is unlikely to invest any great effort in figuring out what the word means.

4. *Helpfulness of surrounding context in understanding the meaning of the unknown word*: a given cue can be differentially helpful depending upon the nature of the word whose meaning is to be inferred and upon the location of the cue in the text relative to the word whose meaning is to be inferred. Consider first an example of how the nature of the word can affect cue helpfulness. A temporal cue describing a *diurnal* event would probably be more helpful than a spatial cue in aiding an individual to figure out that diurnal means daily. In contrast, a spatial cue would probably be more helpful than a temporal cue in figuring out that *ing* is a low-lying pasture. It is unrealistic to expect a given kind of cue to be equally helpful in figuring out the meanings of all kinds of words. Consider, for example, how the location of the cue relative to the unknown word can affect cue helpfulness. If a given cue occurs in close proximity to the unknown word, then it is likely that the cue will be recognized as relevant to inferring the unknown word's meaning. If the cue is separated from the unknown word by a substantial portion of text, the relevance of the cue may never be recognized; indeed, the cue may be misinterpreted as relevant to an unknown word to which it is more proximal. The helpfulness of a context cue may also be mediated by whether the cue comes before or after the unknown word.

5. *Density of unknown words*: if a reader is confronted with a high density of previously unknown words, he or she may be overwhelmed and be unwilling or unable to use available cues to best advantage. When the density of unknown words is high, it can be difficult to discern which of the cues that are available apply to which of the words that are unknown. In such a situation,

utilization of a given cue may depend upon figuring out the meaning of some other unknown word, in which case the usefulness of that cue (and very likely of other cues as well) is decreased.

6. *Usefulness of previously known information in cue utilization*: inevitably, the usefulness of a cue will depend on the extent to which past knowledge can be brought to bear upon the cue and its relation to the unknown word. The usefulness of prior information will in large part depend on a given individual's ability to retrieve information, to recognize its relevance, and then to apply it appropriately.

For example, the proximity of the context cues to the words acapnotic and phalacrosis makes these cues easier to apply. The necessity of understanding the meanings of these two words to understanding the passage as a whole (which is about why the blind date failed) increases reader motivation to learn the words' meanings. One's knowledge that some people are vigorous anti-smokers and that some men, myself excluded, go bald with age, facilitates one's decontextualization of the meanings of the two words. . . . And so on. Our data show that it is sometimes easier to learn a new word for a new concept rather than to use an old word that already has another meaning for a new concept. The latter case promotes negative transfer (Tetewsky & Sternberg, 1986).

To summarize, I have claimed in this part of the chapter that three basic facts about vocabulary—its size, its relation to reading comprehension, and its relation to intelligence—can be explained, at least in part, by a tripartite theory of vocabulary learning.

IMPLICATIONS OF THE FACTS AND FRAMEWORK FOR THE TEACHING AND LEARNING OF VOCABULARY

What are the implications of the facts and framework for the learning and teaching of vocabulary?

Before discussing these implications, I must make clear one thing that the facts and framework do not imply, namely, that learning from context is the fastest or most efficient way of learning specific vocabulary. The inference does not follow logically, and it is not correct empirically. The naturalness or typical use of a method of learning does not imply its optimality. To the contrary, there may be more efficient ways of learning. During the past several years, Joel Levin, Michael Pressley, and their colleagues have performed a series of studies that have convinced me, at least, that for learning specific vocabulary, the keyword method of vocabulary teaching and learning is

faster and more efficient than learning from context (see, e.g., Levin, McCormick, Miller, Berry, & Pressley, 1982; Pressley, Levin, Kuiper, Bryant, & Michener, 1982; Pressley, Levin, & McDaniel, this volume). As far as I can tell, it may be the most effective of the currently available methods. If I have any concerns about that method, they are not with its efficacy, but with the nature and durability of its use. I seek here not to criticize the keyword method, but rather to clarify the conditions of its use.

First, the keyword method, and other methods related to it, require one to know in advance the meaning of the unknown word(s) one wishes to learn. Learning from context does not. If one has definitions available, then learning from context is not so efficient. Why waste a good definition? But during the course of most vocabulary learning, one does not have definitions readily available. Usually, a dictionary is not available. Even if it was, many people are too lazy or busy to look up meanings of the new words they encounter. The point to be made is that during the large majority of word-learning experiences, learning from context is the only feasible method of word learning. If one teaches methods of learning from context, it should be with the goal of teaching vocabulary-building skills, rather than of teaching specific vocabulary.

Second, the keyword method lacks a certain spontaneity. In some respects, the method may be reminiscent of speed reading. Speed reading is often more efficient than normal reading. But how many people continue to use it, well beyond the time they learned to use it? Indications are that the overwhelming majority do not. I could swear that once, after taking the Evelyn Wood Speed Reading Dynamics Course, I read several thousand words a minute. Within a short time, I was back to several hundred. Speed reading requires a great deal of mental effort—more than most people are willing to expend during the extended course of their everyday reading—and it is also, at least for some, uncomfortable. I cannot help but wonder how many people who have learned vocabulary by the keyword method, whether in an experiment or in actual language instruction, will continue diligently to use the method on their own. Thus, my concern in this chapter is not with what people might be able to do, but with what they actually do.

Let me turn now to what the facts and framework do imply. I think they imply that learning from context is central to most everyday vocabulary learning, and that teaching people the processes, cues, and moderating variables of contextual learning can help those people if they do not already use these vocabulary-learning aids spontaneously. In order for these aids to be effective, first, they do have to be used spontaneously. Otherwise, laziness or busyness will intervene to render them as ineffective as other methods of vocabulary building might be in the course of one's everyday life. Second, they have to be used almost automatically. By adulthood, good readers should ap-

ply the processes of selective encoding, selective combination, and selective comparison to available cues while being at most minimally aware that they are doing so.

If one is to teach learning from context, then I believe there are three principles one must keep in mind. Without them, I doubt such teaching will be effective.

First, presentation of words in context is not enough: One needs theory-based instruction in how to use context. In typical contextual training programs, including those used in experimental comparisons to other methods of vocabulary teaching, the learning-from-context method consists of presenting words embedded in a series of sentences, either with or without prior definitions. But if children (or adults) do not know the processes, cues, and moderating variables that can be used for contextual learning, they will not benefit optimally from such instruction. In essence, they have to learn how to learn from context before they actually can learn from it.

Second, presentation of the words in context, plus decontextualization knowledge and skills, is not enough either: One needs relevance of the training to students' everyday lives, and the students must be convinced of this relevance. It took an "ignorant" graduate student to teach me this. A visiting potential graduate student indicated to me and to some of my graduate students that he was interested in the possibility of working with me in my research on context effects. I took this to mean that he was interested in the work I was doing relating real-world environmental contexts to intelligence. My graduate students took this to mean that he was interested in the work I was doing on learning vocabulary from context. Eventually, it occurred to me that it was odd that I was studying two kinds of context, but treating them as totally distinct. It had never even occurred to me that I was regularly using the same word, "context," in two contexts that I was perceiving as totally distinct.

The visiting graduate student has led me to question whether they are so distinct. If one presents students or subjects with uninteresting words in uninteresting contexts, whether for experimental credit or for a test the next day, it is perhaps unreasonable to expect strikingly high levels of learning. In most of one's life, one learns because one wants to or because one truly has to, or both. The kinds of training we have given in our experiments or in certain training programs scarcely constitutes such learning. For many students, it will be quite irrelevant to their lives. Why is context so effective in everyday life, if not in the lab, then? Because context, whether verbal or otherwise, forms the milieu in which we live. We need to learn from it to survive, or to survive well. Our laboratory experiments and training programs scarcely reproduce these kinds of conditions, at least for most people. I am reminded of my experience in reading a particularly difficult journal article. When its author sent it to me, I read it and hardly understood a thing. When I received it to review, I read it and understood it quite well. The second exposure may

have helped. But what helped more was that I needed to understand it the second time, so I made sure I did. In short, then, the ignorant graduate student may have known more than I did.

Third and finally, it is critical in teaching vocabulary to teach students to teach themselves. No matter how many words we teach them directly, those words will constitute only a small fraction of the words they will need to know, or that they eventually will require. They truly constitute a drop in the vocabulary bucket. It doesn't really matter a whole lot how many of those few words students learn, or how well they learn them. What matters is how well they will go on learning long after they have exited from our lives, as we have exited from theirs. This will be true whether we see them during the course of a 1-hour experiment, or a 1-month training program, or whatever. If, indeed, most vocabulary is learned from context, then what we most need to do is not to teach vocabulary from context, but to teach students to use context to teach themselves.

DATA TESTING THE THEORY OF LEARNING VOCABULARY FROM CONTEXT

We have collected data that test both the theory of learning from context and its application to vocabulary learning. This research is ongoing, so this chapter by no means presents the final story. But I am able to summarize here what progress we have made so far, emphasizing the data on testing the theory as applied to teaching vocabulary decontextualization skills.

Initial Test of the Theory

The theory was first tested (Sternberg & Powell, 1983) by asking 123 high school students to read 32 passages of roughly 125 words in length that contained embedded within them from 1 to 4 extremely low-frequency words. Thirty-seven of the words (all nouns) were used in the passages; each target word could appear from 1 to 4 times, resulting in 71 different presentations altogether. Passages were equally divided among four different writing styles: literary, newspaper, scientific, and historical. An additional sample passage was written in the literary style.

The students' task was to define, as best they could, each of the low-frequency words within each passage (except for multiple occurrences of a single word within a given passage, which required only a single definition). Students were not permitted to look back to earlier passages and definitions in making their current responses.

Qualities of definitions were rated independently by three trained raters. Because mean inter-rater reliability was .92, an average of the three ratings

was used as a definition-goodness score for each word for each subject. These averages were then averaged over subjects to obtain a mean goodness-of-definition rating for each word. The main independent variables were ratings of the number or strength of the occurrences of our contextual cues and moderating variables (with the exact nature of the rating depending upon the independent variable) with respect to their roles in helping in the deciphering of the meaning of each low-frequency word in the passages.

Theory testing was done via multiple regression. We used a stepwise multiple-regression procedure in which we allowed only three variables plus a regression intercept to enter into our final models. The decision to limit the number of variables was made on the basis of our judgment of the degree of refinement of our data, and in the hope of minimizing the risks of capitalization upon chance that inhere in stepwise regression. Because of multicollinearity (correlation among) independent variables, it was not possible to make strong inferences regarding the "true" subsets of variables that were differentially relevant from one passage style to the next. Variables that entered into at least one of four regressions were enablement, stative-descriptive, functional-descriptive, and equivalence cues, plus moderating variables of helpfulness and importance. The correlations between predicted and observed goodness ratings were .92 for literary passages, .74 for newspaper passages, .85 for science passages, and .77 for history passages. All of these values differed significantly from zero.

We concluded on the basis of these data that the contextual cues and moderating variables proposed by our subtheories provided good prediction of the goodness-of-definition data, although we certainly do not believe that our model accounted for all of the reliable variance. Indeed, the square roots of the internal-consistency reliability coefficients (based on all possible split halves of subjects) for our four data sets, which place an upper limit on each value of R (the multiple correlation), were all .98 or above, showing that there was considerable reliable variance not accounted for by the fitted model. Nevertheless, fits of the model subsets seemed sufficiently high to merit some optimism regarding our initial attempts to understand differential word difficulty in learning from context. Moreover, performance on the task was successful in distinguishing high from low verbal subjects: Definition goodness ratings for individual subjects correlated .62 with IQ, .56 with vocabulary, and .65 with reading comprehension scores. The data, although extremely limited, are consistent with the notion that the proposed theory of cognitive competence is on the right track, at least in the domain of verbal declarative knowledge.

Instructional Experiments

In a second experiment conducted in collaboration with Elizabeth Neuse, we tested 81 sophomores and juniors in an inner-city high school. The sub-

jects were divided into two basic groups, a training group (59 subjects) and a control (no-training) group (22 subjects). The mean IQ of the subjects was 97, with a standard deviation of 11.

The experimental design in this experiment involved seven independent variables: (a) training group (experimental, control), (b) testing time (pretest, posttest), (c) test format (blank, nonword), (d) cue type (stative descriptive, functional descriptive, class membership), (e) unknown word type (abstract, concrete), (f) restrictiveness of context with respect to the meaning of the unknown word (low, high), and (g) sentence function of the unknown word (subject, predicate). These variables were completely crossed with respect to each other. Treatment group was a between-subjects variable; all other variables were within-subject and were manipulated via a crossed-variables testing arrangement. Two different test forms were used, and half the subjects received the first form as a pretest and the second form as a posttest; the other half of the subjects received the reverse arrangement. Test items, involving either neologisms or blanks (cloze procedure), were each presented in the context of a single sentence. There were 48 items on each test. Scores on the pretest were correlated .74 with an IQ test (Henmon–Nelson) given before training, and .71 with an alternative form of the test given after training. Scores on the posttest were correlated .65 and .64, respectively, with the two administrations of the IQ test.

The training sequence was spread out over six sessions. The seven topics covered were:

1. What is context? Subjects were introduced to the concept of context, and of how context can be used to figure out meanings of words whose meanings are unfamiliar.

2. Using sentence context. Subjects were shown examples of how context can be used in different ways to figure out word meanings, and were invited to try out some examples on their own.

3. 20 questions (spotting cue types). Subjects were motivated to learn about cue types by being given examples of unknown words used in context, and then generating the questions one would want to ask about these words in order to figure out what they meant. In this way, they were informally introduced to the various cue types.

4. Cues I (temporal, spatial, stative-descriptive, equivalence). Subjects were instructed as to four of the kinds of cues, learning what each one is, being given examples of it, and then being asked to use it to figure out meanings of unknown words in context.

5. Using paraphrase to figure out word meanings. Subjects were shown how attempting to paraphrase the meaning of a sentence—saying to oneself what the sentence is trying to communicate—can help one figure out the meanings of unknown words.

6. Cues II (functional-descriptive, causal). Subjects were instructed in two more kinds of cues, learning what each one is, being given examples of it, and then being asked to use it to figure out meanings of unknown words in context.

7. Mystery words (neologisms presented in sentences or paragraphs). Subjects were given fairly lengthy sentences and paragraphs with unknown words, and asked to apply all that they had learned to figure out the meanings of these words.

The six class periods proved ample to cover this range of theory-based material.

In the experimental group, significant main effects were obtained for testing time (posttest higher than pretest), cue type (stative–descriptive hardest, functional–descriptive in-between, category membership easiest), context restrictiveness (higher restrictive more difficult than lower restrictive), and sentence function (predicates harder than subjects). In the control group, significant main effects were obtained for cue type (same ordering of means as previously given) and restrictiveness of context (same ordering of means as previously given). Thus, there was a significant pre- to posttest gain in the trained group, but not in the untrained group. However, the interaction between group and training effect was not statistically significant. In addition, there were a number of statistically significant interactions between independent variables, suggesting that model effects were not wholly independent and additive, but were interactive with each other.

Taken as a whole, these results suggest (a) that subsets of the cues and moderating variables do have additive effects that can be quantified and isolated, (b) that the additive effects are supplemented by interactive ones, and (c) that at least some training of decontextualization skills is possible. The set of results is thus supportive of the ideas in the theory of verbal decontextualization, but emphasizes the need to consider interactions as well as main effects in analyses of model fits.

The greatest disappointment in this experiment was the weakness of the training effects. I believed that enough had been learned from this experiment to make possible the design of an experiment that explicitly looked at training effects, without the "distraction" of also testing other things. The next experiment was therefore directed specifically at obtaining improvements in decontextualization ability through theoretically based training.

In this experiment, 150 New Haven area adults (nonstudents) of roughly average intelligence were divided into one of five conditions. There were three training conditions (with a fourth, combined-procedure condition yet to be run) and two control conditions. Subjects in all three training conditions and one of the control conditions received exactly the same practice words and passages, but differed in the instruction they received (if any) regarding the passages. Passages were similar to those in the first experiment.

The 30 subjects in each of these conditions were given a 25-item pretest and a 25-item posttest measuring skill in figuring out word meanings, as well as other tests. The pretest and posttest were transfer tests, in that they measured skill in figuring out word meanings: They did not merely test recall of words in the practice materials. Our goal was not to train specific vocabulary, but rather to train vocabulary-building skills. All words in the experiment were extremely rare English-language words. The same pretest and posttest words were used in each condition, and training words were the same across conditions. Items were scored on a 0–2 point scale, for a maximum score of 50 points per test. Each training session lasted 45 minutes, exclusive of the various kinds of testing, which brought session length to 2½ hours. The conditions, which were between-subjects, were as follows.

In a *process-training condition,* subjects were taught and given practice using the mental processes (selective encoding, selective combination, selective comparison) alleged by the theory to be involved in figuring out meanings of new words from context.

For example, selective encoding was described as the process of sifting out relevant from irrelevant information. Students were presented with several lines of text that contained a rare word, and were given a detailed explanation of how selective encoding could be used to discover the relevant information about the word's meaning in the text. For example, one of the texts read:

> He first saw a *macropodida* during a trip to Australia. He had just arrived from a business trip to India, and he was exhausted.

Processing of these sentences was described as follows:

> There is much information to weed out. For example, we need not know that the man in the passage was on a business trip. We do not need to know that he had been to India or that his *first* exposure to *macropodida* occurred in Australia. That he saw a *macropodida* in Australia is relevant, but we could care less whether he had seen one a hundred times before in every zoo imaginable. In the first sentence there are two important cues: (1) the man saw a *macropodida,* so *macropodida* must be visible, and (2) *macropodida* can be seen in Australia. The second sentence does not contain any relevant information; the facts of this sentence relate to the man, not to the unknown word.

Practice exercises were then presented in which subjects were asked to underline portions of the text that seemed relevant to the meaning of the unknown word.

In a *contextual-cue training condition,* subjects were taught and given practice using the contextual cues upon which the mental processes operate (e.g., class membership, stative–descriptive).

For example, for Cues I (temporal, spatial, stative–descriptive, and equivalence), setting cues were described as temporal, spatial, and situational in-

formation about contexts in which an unknown word may be found. Examples of each type of cue were given. Subjects were told to judge cues by sorting out information specifically relevant to the unknown word. For example, one of the texts began:

> Two ill-dressed people, the one a haggard woman of middle years and the other a lean young man, sat around a fire where the common meal was almost ready. The mother, Tanith, peered at her son through the *oam* of the bubbling stew.

Cues in the text were explained as follows:

> There are three setting cues in the first sentence, and the phrase, 'around a fire where the common meal was almost ready', contains them all. 'Around a fire' is a spatial cue because it tells where the action takes place. The words, 'common meal', provide a situational cue, and 'almost ready' is an important temporal cue.

Subjects were then given practice in identifying setting cues.

In a *moderating-variable training condition*, subjects were taught and given practice using the moderating variables that affect how well the processes can be applied to the cues (e.g., the location of a cue in the passage relative to the unknown word).

For example, for number of occurrences of an unknown word, subjects were told that multiple occurrences signal a word's importance to a text and provide additional information about its meaning, but they also require integration of information from the cues surrounding each appearance.

One of the passages used as an example follows:

> The *flitwite* was only one of the judicial remedies available to the justices of the Court of the King's Bench in the 11th century, but it was perhaps the most important. Its frequent use added enormously to the treasury's coffers, and new royal expeditures were often financed by the issuance of an increased number of *flitwites*. However, even the most impartial of justices must have handed them down in multitudes, for the *flitwite* was as much a part of 11th century society as the civil tort is of our own. Medieval men and women related in direct and personal ways; therefore, conflict was likely to take the form of actual fighting. In our litigious culture, the law must often deal with more subtle forms of conflict.

The word's definition is given at the end of the passage: "a fine for fighting." The information from each occurrence of the word was explained as follows:

> We learn from the first sentence that the *flitwite* was an important judicial remedy in the 11th century. The context of the second occurrence tells us that the *flitwite* was important because it 'added enormously to the treasury's coffers'.

> The context of the third use of the term informs us that *flitwites* were an important part of 11th century culture. As we read on, we learn that the flitwite appears to be a penalty for fighting.

Subjects were then given practice passages and asked to define target words.

In a *vocabulary-memorization control condition*, subjects were asked to memorize definitions of 75 extremely rare words (that otherwise did not appear in the experiment), and were tested on their memory for these words.

In a *context-practice control condition*, subjects were given exactly the same practice that was given to subjects in the three training conditions, except that the practice occurred in the absence of training.

The mean pretest–posttest gain scores (out of 50 points possible on each test) were 7.2 for the process condition, 5.2 for the contextual-cue condition, 7.6 for the moderating-variable condition, 1.1 for the word-memorization control condition, and 2.6 for the context-practice condition. The results are clear: The training groups showed significantly greater gains than did the control groups. Two additional features of the means are worthy of note: First, as would be expected, the controls receiving relevant practice showed greater gain than did the controls receiving irrelevant memorization. The practice control condition is actually similar to many contextual training programs, which consist of little more than practice. Second, to the extent that other programs involve any training at all, it is in contextual cues, which provide the least facilitation of all three training conditions.

In sum, theoretically motivated instruction in learning words from context can make a significant and substantial difference in people's ability to learn word meanings from context. In just 45 minutes of training, substantial gains in decontextualization ability were obtained. Of course, the durability of this training has yet to be shown.

CONCLUSIONS

Most vocabulary is acquired from context. If one's goal is to develop vocabulary-learning skills, instruction in aspects of decontextualization of word meanings—and particularly in processes of knowledge acquisition, contextual cues, and moderating variables—is an effective way to foster this development. In our own research with a general adult population, training in any one of the three aspects of decontextualization resulted in significant and substantial gains in subjects' ability to figure out meanings of new words, relative to control groups. Thus, training of decontextualization skills develops the natural tendency to use such skills, and potentially enables the individual to apply these skills in his or her everyday life. But there are some caveats on the use of decontextualization training that have perhaps been ignored in much past research on learning vocabulary from context.

First, teaching students specific vocabulary words through the use of context is probably not going to result in gains that are as rapid or as large as the gains attained through the use of other methods, such as the keyword method or possibly even rote instruction. The learning-from-context method is at its best for teaching learning-to-learn skills, not for teaching specific vocabulary. Thus, the purpose of teaching decontextualization skills is, ideally, not the same as teaching the keyword method or teaching vocabulary by rote. As a result, comparisons among the methods may be somewhat misleading. If a school has a given amount of time for teaching verbal comprehension skills, though, I would argue that teaching decontextualization skills constitutes a better use of that time than does teaching specific vocabulary. At best, the specific vocabulary taught will be only a small drop in a very large bucket. Teaching decontextualization skills will enable the student later to acquire new words on his or her own. Thus, one is equipping the student with a transfer mechanism that cannot be gained simply from learning specific vocabulary. Recall, for example, that the tests in our learning-from-context experiments are actually transfer tests. They do not measure merely whether specific vocabulary was learned from context, but whether generalized skills for acquiring new words were so learned and can be applied.

Second, training decontextualization skills is probably appropriate for secondary-school students and a general adult population, but not for highly verbal adult subjects, for example, bright college students. We initially attempted to apply our decontextualization training program to Yale undergraduates, who are already preselected for verbal ability. The results were disappointing, for two reasons. First, students were already performing at near-ceiling levels on our tests, largely because they already had skills either similar or identical to those we were trying to train. Second, their application of these skills was already automatic, and our bringing them back to a controlled, conscious level interfered with rather than facilitated their use of decontextualization. Once information processing is automatized, training in that or related forms of information processing can actually interfere with rather than enhance performance (Wagner & Sternberg, 1984).

Third, a training program for vocabulary-building skills must motivate the students by showing the relevance of the skills being taught to their everyday lives. Our own experience indicates that unless the program motivates the students, it will not succeed.

Finally, such a training program should be theoretically based. We believe that our final program for training vocabulary-building skills was successful because it motivated students, but also because it was based on a theory of decontextualization, rather than merely teaching students by example. In typical programs training specific vocabulary from context, students are expected already to know how to use context to increase their vocabulary. This expectation is unrealistic for many students. Many students need to be taught

how to use context, and what they are taught should be based upon psychological theory rather than upon intuitions that may or may not have any psychological grounding.

To conclude, it has been shown that a psychologically based theory of how people learn vocabulary from context can be useful in training vocabulary-building skills. The validation of the theory stems both from empirical data directly testing the theory and data showing the usefulness of the theory for training vocabulary-building skills. It is tempting, in vocabulary-building programs, to train specific vocabulary, just as it is tempting in mathematics programs to train students in particular mathematical formulas and facts. But ultimately, it is what can be used and transferred that matters most, and my hunch is that decontextualization skills are among the most valuable kinds of knowledge one can have. I am not arguing against training specific vocabulary. But I believe such training ought to be supplemented by training in vocabulary-building skills. In this way, the student can teach him or herself well after any particular course has come to an end.

ACKNOWLEDGMENT

REFERENCES

Levin, J. R., McCormick, C. B., Miller, G. E., Berry, J. K., & Pressley, M. (1982). Mnemonic versus nonmnemonic vocabulary-learning strategies for children. American Educational Research Journal, 19, 121.136.

Pressley, M., Levin, J. R., Kuiper, N. A., Bryant, S. L., & Michener, S. (1982). Mnemonic versus nonmnemonic vocabulary-learning strategies: Additional comparisons. *Journal of Educational Psychology, 74*, 693–707.

Sternberg, R. J., & Powell, J. S. (1983). Comprehending verbal comprehension. *American Psychologist, 38*, 878–893.

Tetewsky, S. J., & Sternberg, R. J. (1986). Conceptual and lexical determinants of nonentrenched thinking. *Journal of Memory and Language, 25*, 202–225.

Wagner, R. K., & Sternberg, R. J. (1984). Alternative conceptions of intelligence and their implications for education. *Review of Educational Research, 54*, 179–223.