In R. Carter + M. McCarthy (eds)
Vocabulary and Language Teaching
Longman, London (1988)

Vocabulary and reading

Paul Nation and James Coady

The purpose of this paper is to discuss the relationship of vocabulary to reading with an emphasis on reviewing the relevant research relating to guessing as well as learning vocabulary in context. The effect of vocabulary on readability is also discussed. Although the focus is on learners of English as a foreign language, research with native speakers provides the main source of information.

The paper begins by looking at the effects of vocabulary knowledge on reading and then looks at how reading increases vocabulary knowledge. A strategy for teaching the guessing skill is proposed and the steps are elaborated with reference to research.

Vocabulary and text readability

In measures of readability of a text, vocabulary difficulty has consistently been found to be the most significant predictor of overall readability (Chall 1958; Klare 1974). Moreover, 'once a vocabulary measure is included in a prediction formula, sentence structure does not add very much to the prediction' (Chall 1958, p. 157). Vocabulary difficulty is estimated in various ways; the most usual are word frequency and/or familiarity and word length. That is, sentences are more readable if they contain words that are of high frequency in occurrence and that are shorter rather than longer. Other measures are the degree to which a word calls up other words quickly – association value – and concrete versus abstractness. Klare (1963) points out that 'The characteristic of words most often measured in readability studies is, directly or indirectly, that of frequency' (p. 167).

However, it must be kept very clearly in mind that readability formulae or predictors are an index or measure of text difficulty, not a causal analysis of why a given text is difficult. That is to say, there are a number of factors in a text which contribute to its ease or difficulty for a given reader, but we can most accurately predict that

fact by measuring one variable, vocabulary, and extrapolating from it to the overall case.

Davis (1968, 1972) did extensive investigation into the question of whether there are identifiable subskills within the overall ability to read. He did empirical correlational studies and factor analysis arriving at four clear factors:

1. recalling word meaning
2. determining meaning from context
3. finding answers to explicit questions
4. drawing inferences

Of all the factors, vocabulary was the most important and had the strongest effect. In subsequent studies by Spearritt (1972) and Thorndike (1973) remembering word meanings was the only consistent subskill which persisted across the various analyses.

Thus, vocabulary knowledge would seem to be the most clearly identifiable subcomponent of the ability to read, at least when one uses current experimental and statistical methodology as the tool of investigation. Yap (1979) concludes that 'causal links probably do exist between vocabulary and comprehension and that vocabulary is likely to be the predominant causal factor' (p. 58).

The effect of low frequency vocabulary

While research indicates that the presence of low frequency vocabulary in a text has a negative effect on comprehension (Marks, Doctorow and Wittrock 1974; Kameenui, Carnine and Freschi 1982; Freebody and Anderson 1983), the answers to the following questions have been difficult to find.

1. *What is the optimal ratio of unknown to known words in a text?*

Marks *et al.* (1974) found that replacing 15 per cent of the words in a reading text with low frequency words led to a significant decrease in comprehension. Freebody and Anderson (1983), however, have called Marks *et al.*'s criteria for high and low frequency words into question. Freebody and Anderson compared two low frequency word ratios – one low frequency word in three content words, and one low frequency word in six content words. Counting both function and content words, these translate into ratios of roughly 1 in 6 (17 per cent) and 1 in 12 (8 per cent). Although there was some decrease in comprehension at the 1 in 12 ratio, it was only at the 1 in 6 ratio that there was a reliable decrease in comprehension. Kameenui *et al.* (1982) found that ratios around one low frequency word in fourteen running words (7 per cent) gave a reliable decrease in correctly

answering inferential questions based on the text. The answering of literal questions was not significantly affected.

Research with foreign-language learners has not provided an answer to the ratio question. Holley (1973) tried to find the best ratio experimentally. She investigated the relationship between new word density (i.e. the ratio of unknown words to the total length of a text) on the one hand, and vocabulary learning, reading time, comprehension, and student rating of difficulty and enjoyability on the other, using a 750-word text with a glossary. Instead of finding a favourable new word density beyond which learning suffered, Holley found that 'vocabulary learning continues to increase even up to a new vocabulary density of one new word per fifteen known words' (7 per cent) (p. 343). Scores on reading time, comprehension, and student ratings of difficulty and enjoyment were not significantly related to new word density.

A reason for Holley's finding may be that her text was short, 750 words, compared with the length of most simplified reading books which are several thousand words long. In Holley's short text, a high ratio of unknown words to known may be acceptable because the total number of unknown words is not high. In a longer simplified reading book, this high ratio would result in an unacceptably high total number of unknown words.

It is likely that only a study involving a large amount of material and a representative range of prose types will provide useful answers to the question of unknown word density. Until there is further research it is still wise to follow the guideline suggested by West (1941, p. 21) of a ratio of no more than one unknown word to fifty known words (2 per cent).

2. *In what ways do low frequency words affect comprehension?*

Freebody and Anderson (1983) examined the effect of placing low frequency words in the important parts of the text as well as in the unimportant parts. The effect of putting difficult vocabulary in important parts of the text was not clear, but seemed to result in a general drop in comprehension over the whole text. The effect of difficult vocabulary in unimportant parts of the text resulted in more adult-like summaries. A 'parsimonious explanation of this result is that students did not process many of the unimportant items, lightening the load in terms of length, and helping them focus on more important items which would be more useful in the formation of summaries' (p. 35). This indicates that readers' reaction to unknown words may be simply to skip over them if they do not seem to play a crucial role in the text.

The effect of pre-teaching vocabulary

Kameenui *et al.* (1982) in two studies found that pre-teaching vocabulary had a significant effect on comprehension. The pre-teaching involved mastery learning where the meaning of the low frequency word was given and the learner answered questions which used the word in a sentence context. As soon as the teaching was completed the learners sat the comprehension test. In an earlier experiment, Pany, Jenkins and Schreck (1982) found only negligible effects of vocabulary training on reading comprehension. Kameenui *et al.* (1982) looked at the effect of redundant information in the text and suggested that the positive effects of this could mask the effects of vocabulary learning. Stahl (1983) found that two of his three groups showed comprehension gains as a result of vocabulary pre-teaching.

Beck, Perfetti and McKeown (1982), McKeown *et al.* (1983) and Omanson *et al.* (1984) examined the effect of vocabulary teaching using a variety of procedures on reading comprehension. The following conclusions can be drawn from their studies.

1. If vocabulary 'instruction is to influence comprehension it must go beyond establishing accurate responses to words' (McKeown *et al.* 1983, p. 17). It must develop fluency of access to word meaning and must integrate the learned words into existing semantic networks.
2. Such instruction takes considerable time. In the McKeown *et al.* (1983) experiment, 104 words were taught over a five-month period in 75 thirty-minute lessons. About 80 per cent of the words were learned.
3. Repetition of the words affected learning with more repetition having some effect on some learners. The minimum number of repetitions in the study was around ten, and this was enough to have an effect.
4. The pre-teaching of vocabulary has an added effect of increasing the saliency of a word when it is met during reading. This meeting gives 'rise to parallel processing in which the learning context of the instructed words is called to mind, which in turn improves the recall of propositions [in the text] containing the instructed words' (Omanson *et al.* 1984, p. 1266).

The studies on readability and pre-teaching indicate the important role vocabulary knowledge plays in reading. But they also indicate the difficulties in experimentally demonstrating a clear connection between vocabulary manipulations and comprehension. Vocabulary knowledge is only one, though an important one, of many factors that allow readers to get information from texts. If, for particular texts, vocabulary knowledge is insignificant, then a range of strategies and other sources of information is available to compensate for this lack. We will look at one of these strategies in the following section of this paper.

Learning vocabulary through reading

Nagy and Anderson (1984) conclude that 'even the most ruthlessly systematic direct vocabulary instruction could neither account for a significant proportion of all the words children actually learn, nor cover more than a modest proportion of the words they will encounter in school reading materials' (p. 304). Jenkins, Stein and Wysocki (1984) point out that 'learning from context is still a default explanation; evidence that individuals actually learn word meanings from contextual experiences is notably lacking' (p. 769). Indeed the very redundancy or richness of information in a given context which enables a reader to guess an unknown word successfully could also predict that that same reader is less likely to learn the word because he or she was able to comprehend the text without knowing the word.

Coady (1979) has argued that the successful ESL reader employs a psycholinguistic guessing approach (Goodman 1976; Smith 1982). That is to say, the reader samples the clues in the text and reconstructs a mental representation of what he or she thinks the text says. This analysis by synthesis approach to reading has also been described as a top-down model of reading. In contrast to this approach, the more traditional view of reading as decoding of letters into sound and ultimately meaning, is characterized as a bottom-up model. More recent theorizing in schema-theoretic models of reading has claimed that both approaches are integral to reading (Adams 1982).

Typically, ESL learners are poor decoders since their vocabulary knowledge is weak while, at the same time, they are already literate in their mother tongue, and are familiar with top-down processing. Therefore, it becomes important to consider whether our instruction should emphasize top-down or bottom-up processing, as well as an appropriate emphasis on the use of context.

Adams and Huggins (1985) claim that word recognition abilities are the single best class of discriminators between good and poor readers. They investigated the sight vocabulary knowledge of second through fifth graders, and proposed a stage theory of sight word acquisition, wherein at the most sophisticated stage the word is 'securely represented in the reader's visual lexicon' (p. 275), i.e. sight

vocabulary; the second stage comprised words not recognized in isolation but only in context, and finally words not recognized at all. Note that sight vocabulary is quite distinct from listening vocabulary where there is no internal mode of the word in its written form. Perfetti and Lesgold (1977, 1979) have argued that when a reader's efforts at word recognition are especially slow and laboured, short-term memory is so taxed that the reader cannot take full advantage of context. In sum, these researchers are arguing that a good reader has a sufficient command over the language so that words are recognized automatically – sight vocabulary – or recognized in context. Poor readers do not have enough sight vocabulary to take advantage of the context. This would seem to imply that successful instruction of ESL readers will have to take into account their vocabulary knowledge and especially their sight vocabulary.

What is context?

Context can be viewed as morphological, syntactic, and discourse information in a given text which can be classified and described in terms of general features. This is the context within the text. But the reader also has background knowledge of the subject matter of a given text, i.e. the general context. Good readers take advantage of such background knowledge in processing the text, and in creating an expectation about the kind of vocabulary that will occur. Hayes-Roth and Hayes-Roth (1977) and Abramovici (1984) have found that lexical information persists in memory representations of meaning; that is to say, good readers tend to remember the words they encounter as well as their meanings.

In an experiment on the facilitating effect of previous knowledge, Adams (1982) found that giving learners information about the topic of a passage before they read it resulted in significantly higher scores on guessing the meanings of nonsense words in the texts. Learners reading in their mother tongue gained higher scores than those reading in a second language, French.

Learning from context

In the research and literature on guessing words from context, a distinction is often made between getting the meaning of a word from the use of context clues, and the learning or retention of this meaning. Studies on getting the meaning give their attention to the types of clues available in context, learners' success or failure in using available clues, and the effect of training on using clues.

Studies on learning words from context sometimes consider the presence of clues, but are most interested in what has been remembered of a word from meeting it in context. Failure to remember information from context can result from failure to get the meaning or from failure to retain the meaning. It is important to note that studies on learning words from context have not shown the large amounts of learning we might expect, considering the rates at which first-language learners seem to increase their vocabulary. (See Anderson and Shifrin 1980).

Jenkins *et al.* (1984) presented low frequency words in very informative contexts in two, six or ten passages read over several days. 'Half of the unfamiliar words were informally taught before their appearance in the passages. Word meanings were learned from context, and more frequent presentation in context increased learning' (p. 707). However, Jenkins *et al.* were surprised that the amount of learning from context was not as great as was expected. Pre-exposure to some of the words by seeing them listed on a sheet with synonyms and a sentence context had a marked effect on learning from context.

Nagy, Herman and Anderson (1985) argue that the failure of Jenkins *et al.*'s study to show substantial learning from context results from the experimenters' failure to consider truly the incremental nature of learning from context. As a result their measures of word knowledge were not sensitive enough to reveal small increments of learning. In their study, Nagy *et al.* used multiple-choice and interview measures which were designed to show small amounts of learning if they occurred. As a result of their research, Nagy *et al.* estimated the probabilities of learning a word from context after just one exposure to be between .10 and .15. Although this seems low, when it is seen in relation to the hundreds and perhaps thousands of unknown words a learner meets, this could result in learning a substantial number of words. And, of course, repeated exposure to a word should have some incremental but as yet undetermined effect.

The rate of success in guessing

What are the chances of success in guessing from context? Ames's (1966) study gives the clearest indication of this because the many words to be guessed were chosen on a random basis. His doctoral level students successfully guessed 60 per cent of the unknown words. Liu and Nation (1984), working with advanced second-language learners, found that the high proficiency learners guessed

between 85 per cent and 100 per cent of the unknown words. The unknown words were all the low frequency words in the texts which were not in *A General Service List* (West 1953), and *A University Word List* (Xue and Nation 1984). The important corollary is that if the learners cited were able to guess a majority of the words, then the necessary clues are there for other, perhaps less gifted, learners to use. It is not an unrealistic goal to expect learners to guess between 60 per cent and 80 per cent of the unknown words in a text if the density of the unknown words is not too high.

Sternberg and Powell (1983) distinguish between clues to the meaning of an unknown word in context, and variables that facilitate or hinder the use of these clues. Density, that is the ratio of unknown to known words in a text, is one such variable. Other variables include the number of times the same unknown word occurs in a text and the variety of contexts in which it occurs in the text, the importance of the unknown word to understanding the context in which it is embedded, the closeness of the contextual information to the unknown word (Carnine, Kameenui and Coyle 1984), and the usefulness of prior knowledge.

A few experiments on training learners to guess from context have shown some improvement in guessing (Hafner 1965, 1967; Carnine *et al.* 1984). Teaching a strategy is one way of providing training.

A strategy for guessing from context

The following strategy is an elaboration of one described by Clarke and Nation (1980). It represents a procedure learners can use to ensure that they are making good use of the available context clues. As will be seen later, it is expected that as the learners become more proficient in the use of the clues, they will not need to follow the steps of the strategy so rigidly.

The strategy presupposes two things; firstly that the learners are able to follow the ideas in the text they are reading, that is, that they have sufficient command of vocabulary, grammar and reading skills in order to achieve basic comprehension of the text, and secondly that the learners bring some relevant background knowledge to the text.

This strategy consists of five steps:

1. Finding the part of speech of the unknown word.
2. Looking at the immediate context of the unknown word and simplifying this context if necessary.
3. Looking at the wider context of the unknown word. This means looking at the relationship between the clause containing the unknown word and surrounding clauses and sentences.
4. Guessing the meaning of the unknown word.
5. Checking that the guess is correct.

Initially the strategy is a major interruption to the reading process while learners develop familiarity with the range of clues available.

Steps 1 and 2: Focusing on the word and its immediate context

The first two steps of the strategy focus on the word itself and the pattern it fits into with the words close to it. Aborn, Rubenstein and Sterling (1959) investigated native-speakers' prediction of words missing from isolated sentences. They concluded that 'increasing the context beyond ten words does not increase predictability. The length at which context attains maximum effectiveness lies between five and ten words' (p. 179). They also found that having context on both sides of a gap was superior to a longer context on either side. If the immediate context is difficult to interpret because of other unknown words, however, then guessing is affected.

Studies of incorrect guesses (Haynes 1984; Laufer and Sim 1985) show that many learners are unable to make use of the immediate context and are often misled by the form of the unknown word.

Step 2, immediate context, can be elaborated by listing possible sources of information that learners can look for:

1. Use the context to answer the question 'What does what?' about the unknown word.
2. Make use of any related phrases or relative clauses.
3. Remove *and* or *or* and make two or more simpler sentences.
4. Interpret punctuation clues such as italics (showing the word will be defined), quotation marks (showing the word has a special meaning), dashes (showing apposition) or brackets (enclosing a definition).

Step 3: Using the wider context

Clauses and sentences in texts enter into relationships with surrounding clauses and sentences. These relationships include cause and effect, contrast, generalization – detail, exclusion (*on the contrary, instead*), explanation (*in other words, that is*), time (*before, subsequently, finally*), and arrangement (*in the first place, secondly*). These relationships may be signalled, but most often they are left for the reader to infer. Helping learners make use of these relationships usually involves making the implicit relationships explicit (Nation 1984).

The wider context can also be elaborated by citing possible sources of information for learners to make use of:

1. Make use of any reference word clues like *this*, *that*, *it*, etc.
2. Complete any comparison clues.
3. Choose and interpret the appropriate conjunction relationships between the clause or sentence with the unknown word and the preceding and following clauses or sentences.

Several researchers have developed lists of the clues which are available in context to help in guessing the meaning of an unknown word. Usually these lists were made to guide teachers in helping their learners develop the guessing skill. The lists were developed in several ways:

a) by analysis of texts (Artley 1943; Dulin 1970);
b) by getting learners to describe the clues they used on words they selected themselves (McCullogh 1943, 1945, 1958);
c) by getting learners to describe the clues they used to guess words which were randomly chosen by the experimenter (Ames 1966).

The lists can be divided into two main types – those based on features of semantics or meaning and those based on sources of clues. Sternberg and Powell's (1983) list is an example of the first type. The list contains eight items and is suited particularly to guessing the meanings of nouns. It acts as a checklist for learners to use to see if the related information is available in the text. Sternberg and Powell suggest that when the learners are trying to guess a word they should look for temporal clues regarding the duration and frequency of the unknown word, value clues, class membership clues, etc. Sternberg and Powell's list describes the type of information to look for, but does not indicate what form that information can take in a text.

The most thoroughly researched list of sources of clues is that produced by Ames (1966), which contains fourteen items. One of these, clues derived from language experience or familiar expressions, does not apply to true guessing from context, because it presupposes that all of the familiar expression is already known. Of the other thirteen items, four can apply to step 2 of the guessing strategy (modifying phrases or clauses, words connected or in series, preposition clues, non-restrictive clauses or appositive phrases), and nine apply to step 3 – the use of wider context. These nine include definition or description, comparison or contrast, synonym, tone, setting and mood, referral, main idea–details, question–answer, and cause–effect.

The aim of most guessing strategies is to make learners aware of the range of information available from context so that after practice they have no need to keep to any rigid guessing procedure.

Step 4: Guessing

Step 4 consists of the actual guess made by the learner using the clues obtained in steps 1–3. This guess may be made in the mother tongue or in English.

Step 5: Checking the guess

There are several ways of checking the guess:

1. Check that the part of speech of the guess is the same as the part of speech of the unknown word.
2. Break the unknown word into parts and see if the meaning of the parts relates to the guess.
3. Substitute the guess for the unknown word. Does it make sense in context?
4. Look in a dictionary.

When the learners have used the available context clues to guess an unknown word, they then can use additional information to check that their guess is correct. The first way of checking is to see if the part of speech of the unknown word is the same as the part of speech of the guess. A surprising number of wrong guesses are a different part of speech from the unknown word. If the learner checks and the part of speech is not the same, then another guess should be made.

A second way of checking is to use the form of the unknown word, particularly prefixes and stem, as a clue to its meaning. For example, *presentiment* can be broken into three parts, the meaning of which can be used to compare with a previous guess of the meaning of the word. It is very important that the use of the word form comes after the context clues have been used. A common source of error with untrained learners is guessing using the form of the word rather than the context (Looby 1939; Gibbons 1940; Haynes 1984; Bensoussan and Laufer 1984). For example, *habitat* was guessed as *habit*, *enormous* as *abnormal*, *offspring* as *the end of spring*, *on the grounds* as *on the earth*, *uniquely* as *unequally*.

When learners make an incorrect guess based on word form, they then try to interpret the context to support the incorrect guess. If they learn to delay using word form clues until after using contextual information, then one of the most difficult parts of the strategy has been mastered.

One important reason why learners rely heavily on the form of the word when guessing is that their vocabulary knowledge is so poor that they cannot interpret the surrounding context (Laufer and Sim

1985). Thus the only source of information they can use is the form of the unknown word. In the Bensoussan and Laufer experiment (1984) many of the learners had to guess at a density of one unknown word in every eleven running words.

However, Laufer and Sim (1985) and Gibbons (1940) showed that even the better readers among their learners made wrong guesses based on form. Haynes's (1984) study clearly shows that second-language learners are likely to let the form of an unknown word take priority over syntactic clues.

Similarly, second-language learners are more adept at making use of syntactic clues than they are at using discourse level clues. Research on reading by Cziko (1978) supports this conclusion. Gibbons (1940), working with university graduates who were native speakers of English, found that 33 per cent (78 out of 234 freshmen) were unable to guess *itinerant* in the following context, and 91 per cent were unable to guess *vicarious*.

> In the beginning the teacher travelled from one locality to another to meet the students, thereby bringing into existence the *itinerant* school master.
>
> Part of our education is obtained directly through actual experiences; *vicarious* experiences which come through reading, pictures, lectures, art and music are equally important, however, as a means of extending real experiences.

Studies of incorrect guessing show the importance of getting learners to delay making use of word clues until they have made full use of the available context clues. For this reason, in a guessing strategy, information based on word part analysis is best used as a way of *checking* context-based guesses. In addition, guessing making the widest use of context clues is encouraged if the context is understandable. If the frequency of unknown words is high, then learners are forced into a word-by-word reading strategy, and they guess by using word form clues rather than context.

Conclusion

In general the research leaves us in little doubt about the importance of vocabulary knowledge for reading, and the value of reading as a means of increasing vocabulary. The precise nature of these relationships, and how we can make use of them in our teaching, are still fruitful areas of investigation.

Points for further development

1. Nation and Coady's review of the research implies that, in spite of a long history of investigation into vocabulary and reading in a first language, and isolated studies in second-language learning, we are still forced to rely on our feelings and intuitions about how we can best deal with vocabulary for reading. In what ways have intuition and experience influenced your approach to the problem of vocabulary for reading? Do you follow any particular set of principles for dealing with vocabulary in the reading lesson?

2. Pre-teaching vocabulary has traditionally been recommended to help learners deal with a reading text. Nation and Coady claim that research indicates that this may be of doubtful value. Firstly, knowing the meaning of a word and readily accessing that meaning both require attention. Secondly, pre-teaching may result in the discouragement of strategies such as guessing, or ignoring unknown words. It may make learners give an importance to knowing the meanings of words in texts which discourages the use of other coping procedures. Research by Taylor (1986) nonetheless suggests that pre-teaching *is* useful and has an important role to play. Do research experiments still have a value for teachers, even when contradictory claims result?

3. There is evidence to show that too high a density of unknown words in a text has a negative effect on comprehension and vocabulary learning. The optimum density is probably a function of a variety of factors not the least being interest in the text. Statistical studies of vocabulary indicate that a relatively small vocabulary is needed to account for a very high percentage of words in a text (Kucera 1982; Nation 1983). If teachers ensure that learners master this important base vocabulary through a variety of approaches, and that reading material is roughly matched to vocabulary level, then comprehension and vocabulary learning activities will have more chance of success. Consider some of the texts you regularly use in your teaching in the light of these remarks.

4. The general conclusion to be drawn from research is that learning vocabulary through context must be the major way of increasing vocabulary knowledge. But it would seem that two complementary approaches are necessary to get this increase: the encouragement of a substantial quantity of reading and the development of the skill of guessing from context. How can we motivate learners who may lack the reading habit to do this 'substantial' reading, and thereby increase their vocabulary?

5. Good learners can guess a very high proportion of unknown words, perhaps 60 per cent to 80 per cent, providing the density of unknown words is not too high. Success in guessing is affected by variables such as the number of times a word occurs, the variety of contexts in which it occurs and the importance of the word in the text. This would still seem to leave the problem of judging the right density of unknown words to the teacher; are there any ways in which teachers can be assisted in this?

6. Nation and Coady suggest a practical strategy for guessing unknown words in texts, consisting of five steps:
a) find the part of speech of the word;
b) examine the immediate context;
c) examine the wider context;
d) guess the meaning;
e) check that the meaning is correct.
Can students be trained to the habit of using the five steps? We might also consider whether some steps are more crucial than others.