Factors Affecting Guessing Vocabulary in Context

Liu Na and **I.S.P. Nation**
Jilin University, China
Victoria University of Wellington, New Zealand

Low frequency words at two different densities in a text were replaced by nonsense words. The subjects were asked to guess the meanings of these words by using context clues. The results were analysed to see what factors affected difficulty of guessing. Words in the low density text (1 unknown word in 25) were easier to guess than words in the high density text (1 word in 10). Verbs were easier to guess than nouns, which were easier than adverbs and adjectives.

The subjects were divided into proficiency levels according to their success at guessing. Groups of learners at high proficiency levels could successfully guess 85% to 100% of the unknown words. The group of learners at the lowest proficiency level tested guessed between 30% and 40% of the unknown words.

From the results, it seems that learning the word guessing skill is best done by the class as a whole rather than individually. At first learners should work on verbs and nouns with plenty of understandable context.

Research needs to be done on the effects of repetition on guessing, the types of words that learners need to guess, and the effect of teaching on improvement of the guessing skill.

Introduction

Guessing the meanings of words from context is the most important strategy for dealing with low frequency vocabulary in written texts. There are many low frequency words and their occurrence is largely unpredictable so it is not possible to learn them in advance. Thus dealing with them as they occur is the only feasible way of handling them. Because of their narrow range and the low probability of meeting them again soon, they do not deserve much effort in learning them. It is better to use context clues to infer their meaning than to spend time on learning the words themselves.

Several writers have described strategies for guessing words from context (Seibert, 1945; Honeyfield, 1977; Clarke and Nation, 1980). This piece of research is an attempt to discover some of the factors that make guessing difficult and to see the potential of the guessing strategy.

Method

The study involved 59 teachers of English as a second language attending a diploma course with proficiency in English ranging from a

few native speakers of English to a few who communicate and read in English with considerable difficulty.

Two sets of passages were used. The appendix contains one set. Each set consisted of a short passage and a long passage which contained a version of the short passage. The passages were made in the following way. All the words in the short passage which were not in the General Service List of English Words (West, 1953) were replaced by nonsense words. These nonsense words were roughly the same length as the words they replaced, were inflected in the same way, and were given suitable suffixes if the original word contained one. So, for example, *dominance* was replaced by *benalment*, *tracts* was replaced by *ranes*. In the short passage about one word in ten had to be replaced. The long passage began with the same original text that was used for the short passage and continued until enough low frequency words were found to match the number found in the short passage. In the long passage, all the words which were not in the General Service List of English Words (West, 1953) and which were not in the American University Word List (Praninskas, 1972) were replaced with nonsense words. Roughly one word in 25 had to be replaced. The same nonsense word forms were used in all passages.

The subjects were given a short passage first and were told to replace all the unknown words with words from their mother tongue or with English words or paraphrases. Most learners chose to answer in English. After they had completed the short passage, it was collected and the long passage was given to them.

Three judges evaluated all the answers to the unknown words in the four passages. The judgement was based on the closeness of the guess to the meaning of the unknown word, rather than its ability to replace the unknown word in the context. For example, for the word *participate*, in the context, "to *participate* in a tutorial discussion", there were the answers *express*, *involve*, *speak* and *contribute*. In this particular sentence, *express* and *speak* make sense. However *participate* means "to have a share" and "taking part". There is a particular sense of involvement in the meaning of the word which neither *speak* nor *express* has. Therefore, *speak* and *express*, both acceptable in the context of this particular sentence, are far from the meaning of *participate* and were marked as wrong answers. According to the degree of closeness in meaning, marks were given of 100%, 80% and 0%. For example, for the word *distinct*, *clear* and *definite* were given 100%; *total*, *real*, *sheer* and *indeed*, 80%; and *big*, *great* and *extreme*, 0%.

Results

The answers were analysed to look at the effects on guessing of the different densities of unknown words and the part of speech of the

unknown words. In addition the success rates of various groups of subjects were examined.

(a) Word density

First, overall scores for the short and long passages were compared. Table 1 contains the results. Each 100% or 80% answer on an item was counted as one correct answer.

Table 1
Group Scores on Two Sets of Passages

Passage	Correct Answers
Origins of Economics (short) n=34	23.4%
Origins of Economics (long) n=24	34.3%
Sally Turner (short) n=25	37.6%
Sally Turner (long) n=24	34.2%

The overall scores on the Origins of Economics passages show a clear superiority of the long (lower density) passage over the short (higher density passage). However this difference is not shown for the Sally Turner passage.

Second, the scores for individual words were examined. In each set the same five unknown words occur in both the long and short passages. Table 2 shows the results.

In all except one case, *militate*, scores for each word in the low density passages are higher than scores in the high density passages. For example 14.7% of the answers for the original word *speculation* were correct in the short (high density) passage. 41.7% of the answers for the same word were correct in the long.

Clearly, different unknown word to known word densities affect guessing from context. The average improvement from a high to a low word density passage ranged from 7% to 10%. Thus, it is best to begin word guessing exercises on low density passages. That is, it is best to make sure that there is plenty of known surrounding context.

Table 2
Scores for the Same Words occurring at Two Densities

Origins of Economics	Short passage	Long passage
Words	Correct answers	Correct answers
speculation	14.7%	41.7%
emergence	20.6%	50.0%
prior	23.5%	29.2%
militate	50.0%	33.3%
scope	20.6%	25.0%
Average	25.9%	35.8%

Sally Turner	Short passage	Long passage
Words	Correct answers	Correct answers
tutorial	64%	70.8%
grade	24%	33.3%
anonymous	12%	12.5%
prospect	16%	25.0%
entitle	64%	75.0%
Average	36%	43.3%

(b) Part of speech

Experiments on learning vocabulary in lists have shown that the part of speech of a word to be learned can affect the difficulty of the item. Rodgers (1969) found that concrete nouns, verbs and adjectives were easier to learn than other parts of speech. In context, verbs and nouns seem to enter into wider ranging relationships than adjectives and adverbs do. For this reason there are more clues to help guess their meanings. Table 3 shows total scores for words grouped according to their part of speech.

The scores show that verbs are the easiest to guess, nouns are next, then adverbs and finally adjectives. The results could have been affected by the small numbers of adverbs and adjectives but it is unlikely that the order would change if their numbers were increased.

Table 3
Ease of Guessing according to Part of Speech

Origins of Economics	n.	Correct answers
verbs	7	47.2%
nouns	14	25.4%
adverbs	2	26.4%
adjectives	4	12.5%

Sally Turner	n.	Correct answers
verbs	9	42.8%
nouns	6	40.8%
adverbs	2	29.2%
adjectives	3	11.3%

The part of speech of words has a marked effect on ease of guessing. Fortunately, verbs and nouns, which are the easiest to guess, are also the most common.

(c) Success rates

Only two words in the passages could not be guessed by anyone. Every other word was guessed, usually by a minimum of two people and often by six or more people. What happens if this guessing activity is done in class? If the class is a group of mixed ability then by working together the learners will be able to guess between 85% and 100% of the words. This means that although one person might not be able to guess 85% of the words, if everyone guesses the ones that they can, then 85% or more of the words will be guessed by the class as a whole. If there are enough clues for one person to guess a word correctly then it should be possible to teach others to recognize and use these clues.

What proportion of the words can be guessed by learners of low proficiency? To answer this question the subjects were divided into four groups according to their score on the guessing test. Table 4 contains the results for one set of passages. Five people scored above 600 on the short passage. This group could guess 85% of the words. Three people scored between 400 and 600. They could guess 71% (10 out of 14) of the words. Eleven people scored between 200 and 400. They guessed 79% of the words. Notice that each of the top three groups guessed over 70% of the words. The lower score of the 400–600 group is partly due to the small

Table 4
Correct Guesses for Different Proficiency Groups

The Origins of Economics (short)

Words	Total score range Above 600 (5)	600-400 (3)	400-200 (11)	Blw.200 (15)	Total of correct guesses
speculation	1	1	3	—	5
environment	2	—	1	1	4
trace	5	3	8	4	20
analytical	—	—	—	—	0
emergence	3	1	2	1	7
prior	5	1	1	1	8
distinct	4	2	7	2	15
militate	3	3	5	6	17
scope	4	1	2	—	7
dominance	3	1	3	2	9
philosophical	—	—	—	—	0
prevailing	2	—	—	—	2
attitude	5	1	2	1	9
acquisition	3	1	3	—	7
Total of words guessed	12 85%	10 71%	11 79%	8 57%	

The Origins of Economics (long)

Words	Total score range Above 600 (5)	600-400 (3)	400-200 (11)	Blw.200 (15)	Total of correct guesses
speculation	5	1	2	2	10
emergence	4	2	4	2	12
prior	5	2	—	—	7
scope	4	1	1	—	6
militated	3	—	3	2	8
aspect	2	1	3	—	6
emerge	5	2	1	1	9
probe	5	4	3	1	13
tract	4	1	1	—	6
discourse	4	2	2	—	8
survive	5	3	3	2	13
consumption	1	1	—	2	4
embed	4	2	—	—	6
Total of words guessed	13 100%	12 92%	10 77%	7 35%	

number of people in the group. The more people there are, the more likely it is that most of the words will be guessed. Fifteen people scored below 200. They guessed 57% (8 out of 14) of the words. There is a tendency

for people in the lower group to be successful in guessing the easier words but they are not restricted to those words. Notice that more difficult words like *environment* and *consumption* were guessed by subjects in the low group. The groups scoring below 200 guessed between 30% and 40% of the words on the other passages.

So, if learners work as a group and share their guesses a high proportion of the words in the passages can be guessed successfully. Thus it is very useful to practise the guessing strategy as a class exercise rather than as an individual exercise. In this way more words are guessed correctly and learners can share information about the clues that they use. As learners improve in handling the strategy, the work can become more individual.

Discussion

There are several difficulties involved in testing the skill of guessing meaning from context. This piece of research has focused on the more formal elements of guessing. However background knowledge of the subject of a passage must play a critical part in successful guessing. Trying to control or measure the effect of this is not easy.

The research would have been much more realistic if the subjects had been guessing words that they really did not know. The use of nonsense words was an attempt to control this. However, if the subjects already knew the word that they were trying to guess they could use clues like typical collocations to help guess the word. Clues like this would not be available to people who really did not know the word before. The word *prior* in the Origins of Economics passage is an example of this. The fact that *prior* is often followed by *to* would be a clue available to a subject who had already learned the word *prior* but would not be available to another subject.

The use of synonyms in English as answers may result in a few more incorrect answers than there should be. For example, for the word *prospect*, there were answers like *fact*, *point*, *idea*, *time* and *situation*, which might be correct answers in the learners' mother tongue when translated. That is, *prospect*, *fact*, *point* etc. can be regarded as synonyms or polysemes of a mother-tongue form. The use of mother-tongue translations as answers would overcome this problem.

The results of this piece of research however do offer some useful guidelines for a teacher. The guessing strategy is best introduced as a class activity, concentrating on verbs and nouns which have plenty of known surrounding context. Once learners master the skill, they should be able to guess at least 85% of the unknown words that they meet. Guessing from context is a very powerful strategy for dealing with low

frequency vocabulary. It deserves a considerable amount of attention in English classes.

Future research

There is still a lot that we do not know about the skill of guessing words from context. In the present piece of research there were not enough repeated words to see the effect of repetition on guessing. In an informal introspective study, one subject on meeting a repetition of a word said, "Oh. I could not guess this word a few sentences back, so it is not worth trying here." Clearly a change in strategy would bring benefits here.

At present some research is under way to see what words learners of English studying at university need to guess. In a study of low frequency vocabulary, Li (1983) found that there were approximately 16 low frequency words per page of a university textbook. Seven of these were technical words closely related to the discipline of study; three were repeated non-technical words; three were non-technical words occurring only once, with useful parts; and three were non-technical words, occurring only once, with no useful parts. It would be useful to know which type of low frequency words have to be guessed.

A great deal of information can be found by getting learners to introspect aloud while they guess from context (Cohen and Hosenfield, 1981). Information from such studies provides valuable indications on where teaching needs to be focused.

Finally, it would be useful to measure the effect of teaching a guessing strategy on improvement of guessing. As learners of English become better at reading English their guessing skill will also improve. Will teaching this skill result in greater improvement?

References

Clarke, D.F. and Nation, I.S.P. 1980. Guessing the meanings of words from context: strategy and techniques. System 8: 3, 211–220.

Cohen, A. and Hosenfield, C. 1981. Some uses of mentalistic data in second language research. Language Learning 31: 2, 285–313.

Honeyfield, J. 1977. Word frequency and the importance of context in vocabulary learning. RELC Journal 8: 2, 35–42.

Li, Aiqun. 1983. Low frequency words in scientific writing. Unpublished paper. Wellington: English Language Institute.

Praninskas, Jean. 1972. American University Word List. London: Longman.

Rodgers, T.S. 1969. On measuring vocabulary difficulty: an analysis of item variables in learning Russian-English vocabulary pairs. IRAL 7, 327–343.

Seibert, L.C. 1945. A study of the practice of guessing word meanings from a context. Modern Language Journal 29, 296–323.

West, Michael. 1953. A General Service List of English Words. London: Longman, Green and Company.

Appendix

I. The high density test

The Origins of Economics

While man's feldination about his material alamution can be prined to ancient times, the development of prinavalic economics is of relatively recent origin. Indeed, the dractment of economics as a separate field of inquiry telad to the Late Renaissance was a meastard impossibility. Everything salerized against it: the nature and limited plout of economic activity, the benalment of the state and church, the force of custom, and the religious and radalosumic beliefs which shaped antelasing alliranes toward human activity for the adianment of wealth. Economic activity for the satisfaction of wants has of course taken place in every age of human history.

II. The low density test

The Origins of Economics

While man's feldination about his material environment can be traced to ancient times, the development of analytical economics is of relatively recent origin. Indeed, the dractment of economics as a separate field of inquiry telad to the late Renaissance was a distinct impossibility. Everything salerized against it: the nature and limited plout of economic activity, the dominance of the state and church, the force of custom, and the religious and philosophical beliefs which shaped prevailing attitudes toward human activity for the acquisition of wealth. Economic activity for the satisfaction of wants has of course taken place in every age of human history. And since man is a thinking animal, it is reasonable to infer that he has always directed some thought to explaining the material alamuts of his life. Feldination on economic matters is therefore undoubtedly as old as human society itself.

Analytical or theoretical economics, however, has a much more recent history. Not until the 18th century did the content of man's feldination about economic phenomena begin to dract as economic analysis

rather than as economic thought. Economics in its preanalytic stages did not exist as a separate subject matter, nor were there analytical tools with which thinkers prined into economic matters. Thus, there are no ranes or emselards we can study to learn about man's earliest feldinations concerning his material environment. The economic ideas of the ancients are largely unsystematic and must be extracted from the writings of the priests, lawgivers, and philosophers. While some of these ideas have antelased into modern times, though usually in altered form, the content of ancient economic thought is little more than a series of unrelated observations and moral pronouncements on production, prinalment, and exchange adianned in writings devoted primarily to religion, ethics, politics, or law.