Mnemotechnics in Second-Language Learning

RICHARD C. ATKINSON *Stanford University*

For some time I have been involved in efforts to develop computer-controlled systems for instruction. One such effort has been a computer-assisted-instruction (CAI) program for teaching reading in the primary grades (Atkinson, 1974) and another for teaching computer science at the college level (Atkinson, in press). The goal has been to use psychological theory to devise optimal instructional procedures—procedures that make moment-by-moment decisions based on the student's unique response history. To help guide some of the theoretical aspects of this work, research has also been done on the restricted but well-defined problem of optimizing the teaching of a foreign language vocabulary. This is an area in which mathematical models provide an accurate description of learning, and these models can be used in conjunction with the methods of control theory to develop precise algorithms for sequencing instruction among vocabulary items. Some of this work has been published, and those who have read about it know that the optimization schemes are quite effective—far more effective than procedures that permit the learner to make his own instructional decisions (Atkinson, 1972a, 1972b; Atkinson & Paulson, 1972).

In conducting these vocabulary learning experiments, I have been struck by the incredible variability in learning rates across subjects. Even Stanford University students, who are a fairly select sample, display impressively large between-subject differences. These differences may reflect differences in fundamental abilities, but it is easy to demonstrate that they also depend on the strategies that subjects bring to bear on the task. Good learners can introspect with ease about a "bag of tricks" for learning vocabulary items, whereas poor learners are incredibly inept when trying to describe what they are doing.

These subject reports, combined with our own intuitions, led Michael Raugh and me to carry out a series of studies on mnemonic techniques for vocabulary learning. Michael Raugh is a computer scientist and mathematician by training, but throughout his life he has been intrigued by mnemonics of one sort or another; he was the one who convinced me that this line of research was worth pursuing.

The Keyword Method

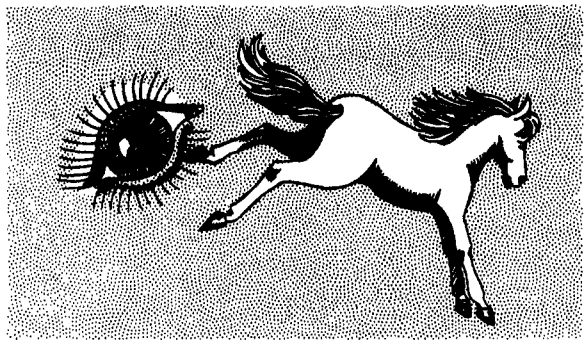
Our initial experiments were not as successful as we had anticipated, but they did help us to develop and refine a mnemonic aid for vocabulary learning that we have dubbed the *keyword method*. It is this method and related experiments that are discussed in this article. By a keyword we mean an English word that sounds like some part of the foreign word. In general, the keyword has no relationship to the foreign word except for the fact that it is similar in sound. The keyword method divides vocabulary learning into two stages. The first stage requires the subject to associate the spoken foreign word with the keyword, an association that is formed quickly because of acoustic similarity. The second stage requires the subject to form a mental image of the keyword "interacting" with the English translation; this stage is comparable to a paired-associate procedure involving the learning of unrelated English words. To summarize, the keyword method can be described as a chain of two links connecting a foreign word to its English translation. The spoken foreign word is linked to the keyword by a similarity in sound (what I call the *acoustic link*), and in turn the keyword is linked to the English translation by a mental image (what I call the *imagery link*).

Let us consider a few examples from Spanish and Russian, the two languages that we have used for most of our research on the keyword method. In Spanish the word *caballo* (pronounced something

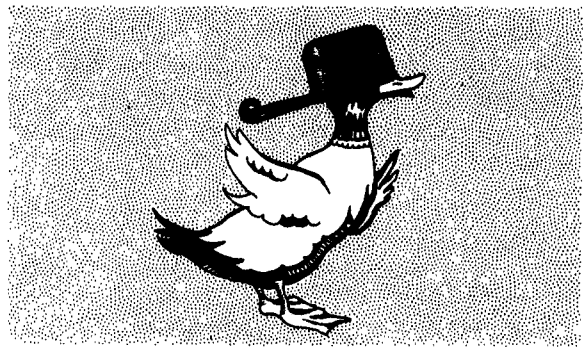
---

This article is based on the presidential address presented to Division 3 at the meeting of the American Psychological Association, New Orleans, September 1974.

Requests for reprints should be sent to Richard C. Atkinson, Department of Psychology, Stanford University, Stanford, California 94305.



CABALLO — eye — HORSE



PATO — pot — DUCK

Figure 1. Two illustrations of how mental images might be used to associate a spoken Spanish word with its English translation.

like "cob-eye-yo") means horse. The pronunciation of the Spanish word contains a sound that resembles the English word *eye*. Employing the English word *eye* as the keyword, one might form a mental image of something like a cyclopean eye winking in the forehead of a horse, or a horse kicking a giant eye. As another example, the Spanish word for duck is *pato* (pronounced something like "pot-o"). Using the English word *pot* as the keyword, one could image a duck hiding under an overturned flower pot with its webbed feet and tufted tail sticking out below.

In Russian the word *zvonók* [1] means bell. Its pronunciation is somewhat like "zvahn-oak," with emphasis on the last syllable, and it contains a sound that resembles the English word *oak*. Employing the English word *oak* as the keyword, one could imagine something like an oak with little brass bells for acorns, or an oak in a belfry, or perhaps an oak growing beneath a giant bell jar. As another example, the Russian word for "building" (*zdánie*) is pronounced somewhat like "zdawn-yeh" with emphasis on the first syllable. Using *dawn* as the keyword, one could imagine the pink light of dawn reflected in the windows of a tall building. Additional Russian examples are given in Table 1.

One procedure for applying the keyword method is to present the subject with a series of foreign words. As each foreign word is pronounced, its keyword and the English translation are displayed. During the presentation of each item the subject must associate the sound of the foreign word to the keyword, and at the same time generate a mental image relating the keyword to the English translation. Because of the similarity in sounds, the acoustic link is formed easily; the imagery link is like learning to associate a pair of unrelated English words by using imagery as a mnemonic aid. One qualification must be added to the above description. The keyword need not always be a single word; for some items it may be a brief phrase if that phrase is particularly salient. What this means for a polysyllabic foreign word is that anything from a monosyllable to a longer word or even a short phrase that spans the whole foreign word might be used as the keyword.

We have been conducting experiments for almost two years on one or another aspect of the keyword

TABLE 1

Sixteen Items from the Russian Vocabulary with Related Keywords

Russian	Keyword	Translation
STRANÁ	[strawman]	COUNTRY
LINKÓR	[Lincoln]	BATTLESHIP
DÉLO	[Jello]	AFFAIR
ZÁPAD	[zap it]	WEST
TOLPÁ	[tell pa]	CROWD
ROT	[rut]	MOUTH
GORÁ	[garage]	MOUNTAIN
DURÁK	[two rocks]	FOOL
ÓSEN'	[ocean]	AUTUMN
SÉVER	[saviour]	NORTH
DYM	[dim]	SMOKE
SELÓ	[seal law]	VILLAGE
GOLOVÁ	[Gulliver]	HEAD
TJÓTJA	[Churchill]	AUNT
PÓEZD	[poised]	TRAIN
CHELOVÉK	[chilly back]	PERSON

[1] Russian words are presented using a standard transliteration of the Cyrillic alphabet into the Roman alphabet; stress is marked.

method. Let me describe in some detail one of the experiments using a Russian vocabulary and then use it as a springboard for discussing other results.[2]

In this experiment, subjects learned a vocabulary of 120 Russian words; the total vocabulary was divided into three comparable 40-word subvocabularies for presentation on separate days. The experiment was run under computer control and involved two independent groups of subjects—a keyword group and a control group. The computer presented prerecorded Russian words through headphones; keyword and English translations were presented on a cathode-ray-tube (CRT) display; and the subject entered his responses into the computer by means of a typewriter keyboard. The experiment began with an introductory session during which subjects were familiarized with the equipment and given some instruction in Russian phonics; subjects in the keyword group were also given instructions on the keyword method. On each of the following three days, one of the subvocabularies was presented for a cycle of three study/test trials. The study part of a trial consisted of a run through the subvocabulary; each Russian word was pronounced three times and simultaneously its English translation was displayed on the CRT. For the keyword subjects, the keyword, set off in brackets, was also displayed on the CRT. The test phase of a trial was exactly the same for both groups. It consisted of a run through the subvocabulary in which each Russian word was pronounced, and the subject had 15 seconds to type the translation; no feedback was given. A comprehensive test covering the entire vocabulary of 120 items was given on the fifth day of the experiment. Without warning, subjects were called back six weeks later for a second comprehensive test.

Figure 2 presents the probability of a correct response over test trials for each of the three instructional sessions. The keyword group in all cases obtained superior scores; in fact, each day the keyword group learned more words in two study trials than the control group did in three trials.

Table 2 gives results for the comprehensive test and for the delayed comprehensive test given six weeks later. Results are presented for the total

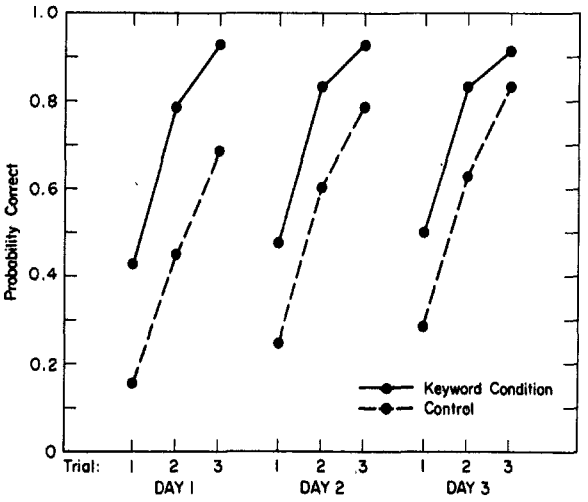


Figure 2. Probability of a correct response over test trials on Day 1, Day 2, and Day 3.

vocabulary and also for the subvocabularies learned during each of the three instruction sessions. I am not commenting on the results as a function of the day on which items were studied—the pattern of the data is what would be expected. The important observation is in the bottom row of the table. Note that for the total vocabulary the keyword group recalled 72% of the items on the comprehensive test, whereas the control group recalled only 46%. Six weeks later the keyword group recalled 43% of the words and the control group 28%. The ratio of control to experimental scores is .64 on the comprehensive test and .65 on the delayed comprehensive test. These are indeed large differences and highly significant statistically.

TABLE 2

Probability of a Correct Response on the Comprehensive and Delayed Comprehensive Tests as a Function of Experimental Treatment and Study Order

Vocabulary	Comprehensive test		Delayed comprehensive test	
	Keyword	Control	Keyword	Control
First subvocabulary	.64	.33	.48	.25
Second subvocabulary	.70	.43	.44	.30
Third subvocabulary	.81	.63	.36	.29
Total vocabulary	.72	.46	.43	.28

[2] An account of this and other experiments using a Russian vocabulary can be found in Atkinson and Raugh (1975).

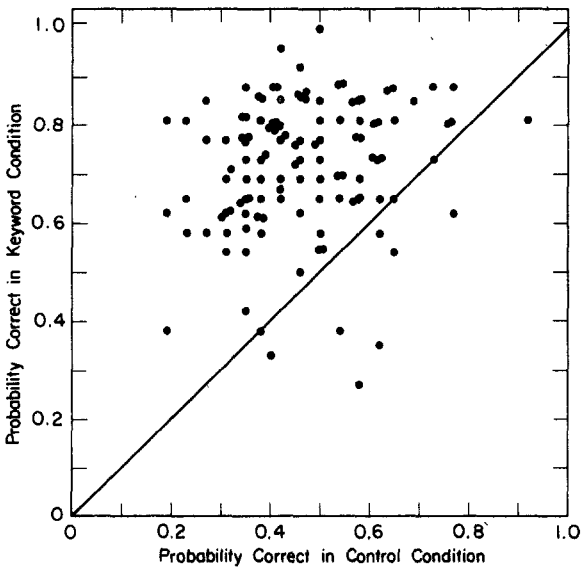


Figure 3. Scatterplot of performance levels on the comprehensive test. (Each point corresponds to an item; the ordinate gives the performance level when the item was studied in the keyword condition, and the abscissa its value when studied in the control condition.)

Figure 3 presents data on the comprehensive test in which each point refers to one of the vocabulary items. The value on the y-axis denotes the keyword group's performance on a word, and the x-axis the control group's performance. Points above the diagonal are for words that were learned best in the keyword condition; below the diagonal are the words learned best in the control condition. Of the 120 words, only 8 were learned better in the control condition. As we shall see later, the poor performance on certain items in the keyword condition can be predicted based on independent estimates of the ease of learning either the acoustic or the imagery link.

This experiment is one in a series of studies that have demonstrated the effectiveness of the keyword method. Our most dramatic demonstration involved a similar experimental design using a Spanish vocabulary.[3] The principal difference was that the control group was told to use a rote rehearsal procedure when studying items. None of the control subjects objected to the rehearsal procedure or found it unnatural, but on a comprehensive test they recalled only 28% of the words. The keyword group recalled 88%. In the Russian experiment described above, the control subjects were highly motivated to do well and were encouraged to use whatever strategies they thought would be most effective. The observed difference between the keyword and control subjects was not a matter of motivation; both groups were highly motivated and very attentive to the task.

Questions about the Keyword Method

Let me now turn to some questions raised by the keyword method and provide what answers I can based on our other experiments.

Question: Should the experimenter supply the keyword or can the subject generate his own more effectively? The answer to this question is somewhat complicated, but, in general, our results indicate that providing the keywords for the subject is best. In a Russian experiment similar to the one described earlier, all subjects were given extensive instruction in the keyword method. During the actual experiment half of the items were presented for study with a keyword, whereas no keyword was provided for the other items. The subjects were instructed to use the keyword method throughout. When a keyword was provided they were to use that word; when no keyword was provided they were to generate their own. On the comprehensive test the subjects were significantly better on the keyword-supplied items than on the others, but the size of the difference did not approximate the difference observed in the last experiment. Instruction in the keyword method is helpful, but somewhat more so if the experimenter also supplies the keywords.

It should be kept in mind that these results are for subjects who have not had previous training in Russian. It may well be that supplying the keyword is most helpful to the beginner and becomes less useful as the subject gains familiarity with the language and the method. We have run another experiment in which subjects were instructed in the keyword method, but during study of an item they received a keyword only if they requested it by pressing an appropriate key on their computer console; we call this variant of the keyword method the *free-choice procedure*. When an item was initially presented for study, a keyword was requested 89% of the time; on subsequent presentations of the item, the subject's likeli-

[3] For a detailed account of this experiment and other work on the keyword method using a Spanish vocabulary, see Raugh and Atkinson (1975).

hood of requesting the keyword depended upon whether he missed the item on the preceding test trial. If he missed it, his likelihood of requesting the keyword was much higher than if he had been able to supply the correct translation. Otherwise, however, the likelihood of requesting a keyword was remarkably constant from one day of the experiment to the next; that is, there was no decrease in keyword requests over the three study days, where on each day the subject learned a new vocabulary. It is interesting to note that performance on the comprehensive test for the free-choice group was virtually identical to the performance of a group that was automatically given a keyword on all trials. Not much of a difference would be expected between the two groups because the free-choice subjects had such a high likelihood of requesting keywords. Nevertheless, these findings suggest that the free-choice mode may be the preferred one. In the free-choice procedure, subjects reported that they generally wanted a keyword but that there were occasional items that seemed to stand out and could be mastered immediately without the aid of a keyword. In summary, the answer to the question is that subjects appear to be less effective when they must generate their own keywords; but results from the free-choice procedure indicate that keywords need only be supplied when requested by the subject.

Question: Does supplying the imagery link for the subject facilitate learning? The answer to this question seems to be no—it is better to have the subject generate his own image linking the keyword to the English translation. We have tried, for example, to supply the imagery link by using cartoonlike drawings and also by presenting brief phrases or sentences linking the keyword and English translation in a meaningful way. Although these experiments were more in the nature of pilot studies, results indicated that subjects performed best when required to generate their own imagery link.

Question: When a foreign word is presented, does the time to retrieve its English translation depend on the method of learning? Unfortunately, I can report on only one study that bears on this issue. A study-test procedure was used in which the subjects alternated between studying the vocabulary list and then being tested on the list. The tests were the same for all items, but two different study procedures were used. For half of the items a keyword was provided during study and subjects were instructed to learn these items by the keyword method; for the other half of the items no keyword was provided and subjects were told to use only rote rehearsal. Subjects were run for a large number of trials. As would be expected, the keyword items were learned at a faster rate than were the rote-rehearsal items, but eventually performance was perfect for both groups. Our interest was in the speed of response. In general, the reaction times correlated very highly with the probability of a correct response, and otherwise did not depend on the method of learning. At asymptote, reaction times were the same for both groups of items. More work needs to be done using response-time measures. We need to extend our experiments over longer periods of time, and also determine if context effects that occur when one is actually trying to use a language influence retrieval processes. These are difficult questions to answer, but available evidence indicates that the method of learning does not affect retrieval times, particularly once an item has been thoroughly mastered.

Question: Are the imagery instructions critical in the keyword method, or can the subject do equally well when told to associate the keyword and English translation by generating a meaningful sentence connecting the two words? In paired-associate learning experiments in which both the stimulus and response are unrelated English words, Anderson and Bower (1973, p. 456) reported that imagery instructions yielded the same results as sentence-generation instructions. Their findings are not in accord with our own. Imagery instructions have a significant advantage over sentence-generation instructions when using the keyword method (73% versus 64%). We are not sure why our results do not accord with those reported by Anderson and Bower, but there are enough differences between the experimental situations that I do not consider it critical. Let me simply note that the imagery instructions are more effective in our situation. The reason, I believe, is a matter of elaboration. After being tested on an item, the subject often realizes that his initial image was not particularly effective; on the next study trial he elaborates and adds details to that image to make it more salient. If a sentence is generated on the initial study and it proves to be a poor mediator, the subject would not be able to elaborate it except by making the sentence longer and more complex—a procedure that intuitively does not seem particularly effective. From this viewpoint, imagery and sentence-generation instructions might be equally good for easy associations,

but imagery would be better as the task becomes more difficult.

Our present instructions for the keyword method in their most general form ask subjects to picture an imaginary interaction between the keyword and the English translation, but we also state that if an image does not come quickly to mind, they may want to try to relate the two words by generating a phrase or sentence. Subjects report that there are items where a sentence pops into mind and seems to be the most natural way of forming the association. The distinction between imagery and sentence generation may be debated on theoretical grounds, but most of our subjects report such differences when introspecting about their thought processes.

Question: How useful is the keyword method if the subject is asked to retrieve the foreign word when given its English translation? We have run one experiment that bears on this question using a Spanish vocabulary. One group of subjects learned by the keyword method and another by rote rehearsal. During learning, subjects studied and were tested on only the forward associations, that is, going from the foreign word to the English translation. All subjects were brought to the same criterion on the forward associations, which, of course, required fewer trials for the keyword group than for the rote-rehearsal group. Immediately thereafter they were tested on the backward associations—they were given the English word and asked to produce the foreign word. Judges blind to the experimental treatments evaluated the responses. On the backward associations the keyword subjects had a score 19% above that of the rote-rehearsal subjects. Even though the forward associations had been learned to the same criterion, the keyword group was significantly better on the backward associations.

Effectiveness of a Keyword

Let me now turn to a somewhat different issue. Data on individual items learned by the keyword method indicate that some keywords are clearly better than others. If an item does particularly poorly in one experiment, its performance can usually be improved in the next experiment by selecting a new keyword. Those familiar with the literature on word variables such as concreteness, imagery, and frequency will not be at a loss for possible explanations. We have examined some likely hypotheses in an article soon to be published, but time does not permit me to review that work here. From a practical viewpoint, the important remark is that keywords should be selected using empirical criteria. When there is not enough time to make empirical determinations, a committee of individuals familiar with the language should select the keywords rather than having one person make the decisions. Experience indicates that individual experimenters can come up with some pretty bizarre keywords that work for them but for no one else. A committee approach seems to protect against this problem.

One empirical procedure for evaluating keywords involves having a group of subjects learn only the foreign-word-to-keyword link, and an independent group learn only the keyword-to-translation link. We have conducted such an experiment with the 120-word Russian vocabulary used in the study that I first described. For each item, an estimation was obtained for the probability of a correct response averaged over the first two test trials. Let me denote that probability as A for the group learning the acoustic link and I for the group learning the imagery link. Finally, let K be the probability of a correct response averaged over the first two test trials for an item in the keyword group in our original experiment. It turns out that the product of $A \times I$ (i.e., Probability of Knowing the Acoustic Link $\times$ Probability of Knowing the Imagery Link) is a fairly good predictor of performance in the keyword condition. Table 3 displays the correlation matrix using rank-order correlations. Note that the correlation between A and I is near zero, indicating that the learning of the acoustic link is not related to the learning of the imagery link. Note also that the correlation between the product $A \times I$ and the variable K is .73; the product is a fairly accurate predictor of performance in the keyword condition. The C entry in the table is comparable to the K entry, except that it denotes performance for the control

TABLE 3

Rank-Order Correlation Matrix for the Variables $A \times I$, K, C, A, and I

Measure	1	2	3	4	5
A × I (1)	1.0	.73	.39	.68	.71
K (2)		1.0	.38	.53	.49
C (3)			1.0	.33	.19
A (4)				1.0	.02
I (5)					1.0

group in our original experiment. Note that C is not nearly as good a predictor of K as is the product of $A \times I$. These results suggest that initial learning in the keyword condition can be interpreted as the learning of two independent links —the acoustic link and the imagery link.

Early in the learning process, the memory structure for a given item involves only these two independent links; however, with continued practice a third link is formed directly associating the foreign word with its English translation. It is this direct link that sustains performance once an item is highly practiced; the subject may still access the keyword, but the retrieval process of the direct association is so rapid that the subject only recalls the keyword under special circumstances, such as when he is consciously trying to do so or when he has a retrieval failure in the main process. But the less direct chain of the acoustic and imagery links has the advantage that it is easily learned and provides a crutch, if you will, for the subject as he learns the direct association—it facilitates the learning of the direct association by insuring that the subject is able to recall items early in the learning process.

Applications of the Keyword Method

In deciding whether to use the keyword method, several problems need to be considered. One problem is that keywords might interfere with correct pronunciation. Our experiments do not deal with this issue, but we have discussed it with experts on language instruction. Opinions vary, but most believe that the keyword method may well facilitate, rather than interfere with, pronunciation. One reason is that the keyword method has features in common with the method of "contrasting minimal pairs"—a standard technique for teaching the phonetics of a foreign language. But even if there were some interference, the keyword method might still be warranted if the rate of vocabulary acquisition was improved substantially.

Another problem to be considered is whether items learned using the keyword method take longer to be recalled. We described some experimental results indicating that asymptotic response times are independent of the method of learning. These results reinforce our experiences with the keyword method. Once an item has been thoroughly learned, it comes to mind immediately, and rarely is the subject aware of the related keyword unless he makes a conscious effort to recall it. What evidence we have indicates that the keyword does not slow down or otherwise interfere with the retrieval process.

Our experimental results convinced us and members of the Slavic Languages Department at Stanford that the keyword method needed to be evaluated in an actual teaching situation. Accordingly, we developed a vocabulary-learning program designed to supplement the second-year course in Russian. The program operates under computer control and follows a procedure similar to one used in our experiments. When a word is presented for study it is pronounced by the computer and simultaneously the English translation is displayed on a CRT. The student is free to study the item any way he pleases, but if he presses a button on his console the keyword is displayed instantly on the CRT. In our first evaluative efforts, students in the second-year course were run on the program for four 40-minute sessions per week over a 10-week period. Each week a new vocabulary of approximately 75 words was presented. The words were classifiable as either nouns, verbs, or adjectives; only the imperfective form of verbs was used. The analyses of these data are still incomplete, but several remarks can be made at this time. First, we experienced no difficulty in selecting keywords for such a large vocabulary and foresee no problems in generating keywords for even larger vocabularies. Second, students were enthusiastic about the procedure throughout the 10 weeks, and in interviews at the end of the program voiced the opinion that the keyword method was very helpful. When the computer program presented a word for the first time, students were likely to request a keyword; the request rate was about 72% during the first week and rose steadily over the 10 weeks to about 83% for the last week. In interviews at the end of the program, students reported that the keyword method worked best for nouns, less well for verbs, and least well for adjectives. However, on a delayed test over the entire vocabulary, subjects did equally well on nouns and verbs, with somewhat poorer performance on adjectives.

We plan to make several improvements in the vocabulary-learning program and to reevaluate it more extensively in the near future. But the first large-scale application of the keyword method proved to be very encouraging and was well received by the Slavic-language faculty. For Russian, more so than many languages, the mastery of a basic vocabulary is incredibly difficult. One of the instructors that we have worked with has

told me that he believes the major obstacle in teaching Russian is not learning the grammar but in mastering a sufficient vocabulary so that a student can engage in meaningful conversations and read materials other than the textbook.

Concluding Remarks

In recent years there has been a revival of interest in mnemonic techniques. Introductory psychology textbooks—mine included—that did not mention the topic a decade ago now give it a great deal of prominence. Some classroom demonstrations of mnemonics are indeed impressive; but beyond impressing one's students, it is difficult to identify instructional situations in which mnemonic aids are truly useful. In Cicero's time these aids may have had some usefulness, but in this age of cheap memory devices (including pencil and paper) the value of mnemonic aids is questionable. An exception may be the keyword method. If our instructional applications prove as successful as the experimental work, then the keyword method and variants thereof deserve a role in language-learning curricula. It may prove useful only in the early stages of learning a language and more so for some languages than others. But there is the promise that the poorer learners receive special benefits, particularly if given some coaching along the way. One limitation of our experiments is that we only provide students with written instructions in the keyword method, and thereafter they are on their own. In an ideal situation, students should be coached by an expert until they are proficient in the skill. Imagine trying to learn tennis by reading a set of instructions and then being left to perfect the skill on your own. Coaching should involve having an expert in the keyword method discuss with the student problems that he may be having, critique the student's images suggesting improvements or alternatives, and in general help perfect the skill.

Our work on the keyword method has not led to any new theoretical insights or even to experiments that have direct relevance to current issues in the psychology of memory. But the research illustrates the steps necessary to take an idea that emerged in the confines of an experimental psychologist's laboratory and develop it to a point at which it can be used in a practical teaching situation.

REFERENCES

Anderson, J. R., & Bower, G. H. *Human associative memory.* Washington, D.C.: Winston, 1973.

Atkinson, R. C. Ingredients for a theory of instruction. *American Psychologist,* 1972, *27,* 921–931. (a)

Atkinson, R. C. Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology,* 1972, *96,* 124–129. (b)

Atkinson, R. C. Teaching children to read using a computer. *American Psychologist,* 1974, *29,* 169–178.

Atkinson, R. C. Adaptive instructional systems: Some attempts to optimize the learning process. In D. Klahr (Ed.), *Cognition and instruction.* Hillsdale, N.J.: Erlbaum, in press.

Atkinson, R. C., & Paulson, J. A. An approach to the psychology of instruction. *Psychological Bulletin,* 1972, *78,* 49–61.

Atkinson, R. C., & Raugh, M. R. An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory,* 1975, *104,* 126–133.

Raugh, M. R., & Atkinson, R. C. A mnemonic method for learning a second-language vocabulary. *Journal of Educational Psychology,* 1975, *67,* 1–16.