ERROR ANALYSIS IN VOCABULARY ACQUISITION

— IN CASE OF AMERICAN CHILDREN —

BY KEIKO KUBOTA

Introduction

It is widely known that the amount of vocabulary knowledge is a good predictor of a person's discourse comprehension [Anderson and Freebody, 1979; Johnson et al., 1978]. But how do children acquire and expand their vocabulary knowledge?

Two kinds of models are distinguished to explain the semantic memory representation. One is a network model and the other is a set-theoretical model (Rips, Shoben, and Smith, 1973). The network model assumes that in semantic memory the network of relations of the words that a person has heard or seen before plays the main role in assigning the meaning. On the other hand, the set-theoretical model stresses that the meaning of a word is represented by the set of semantic components. In explaining the acquisition of new words, these two models seem to be complementary to each other. In general, learning a new thing, including a word, is a "decoding" of its representation. We determine the concept from the relations of already-known concepts. Generative rules are acquired through these concepts.

It is obvious that something is happening in our mind when we see a word and recognize its meaning. Some research has been done to identify the encoding and decoding system or information processing system. Walker (1973) and Rubenstein, Lewis and Rubenstein (1971) reported the effects of pronounceability of words and non-words in encoding and recognition. Fredericksen (1978) explored the effects of visual familiarity on times for identifying words and non-words using a lexical decision task. It was also reported that low ability readers were more susceptible to the effects of visual familiarity. From the linguistic viewpoints, it was pointed out that poor readers lack the "ability to direct attention to the significant visual and auditory stimuli in word recognition situations" (Langman, 1960). In Barganz's research (1971), good readers more ably recognized the underlying forms of words than poor readers. Anderson and Freebody (1980) compared the characteristics of non-words that were affirmed mostly by the high ability students and those affirmed by the low ability students, and reported that they are qualitatively different. They developed different models for the decision sequence for high and low ability groups. Hence, concerning the encoding and decoding system, we could hypothesize that there is a difference between good readers and poor readers. Good readers can focus on the underlying phonemic relationship while poor readers attempt to recode graphemically.

Vocabulary acquisition or a part of it could be rule-governed behavior, if the hypothesis is true. If learning comprises a rule-governed behavior,

(1) "certain categories of learners can be expected to respond in the same way across tasks" and
(2) "learners who have mastered the rule can be expected to perform all tasks in the set correctly" (Bergan et al., 1981).

In the mathematics education field, it is clear that the operations are rule-governed behavior. Approaches to identify the cognitive process by diagnosing students' errors have been taken in that field. Brown and Burton (1978) developed an adaptive testing system to diagnose "bugs" based on the individual cognitive information-process network. Tatsuoka, Birenbaum, Tatsuoka and Baillie (1980) developed a system of "error vectors" which comprise all possible operations and generate rules that students take.

The purpose of present research is to develop a method to identify cognitive operations and strategies in vocabulary acquisition by applying the error analysis approach. The decoding system of subjects will be detected by using a multidimensional binary vector which contains the information on all possible errors students could make when solving a problem as well as that on the general characteristics of the item. Their cognitive and linguistic error pattern will be examined by referring to their response to each item. The multidimensional vector will be made based on the cognitive operation hypothesis when a student is given a task to decide if a presented string of letters is meaningful or not.

The hypothesis of this research is as follows:

Children learn the systems of grammar by breaking down each system into parts and developing rules to combine the parts even when acquiring vocabulary. Therefore, when they decide whether they know the meaning of the stimulus or not, they use the rules they have acquired. There should be some difference between a good learner and a poor learner who are in different stages of vocabulary acquisition.

Method

Subjects

Two forms of single-word tests were administered to fifth grade students in Mahomet, Illinois, in 1978. The responses of the students who took both tests will be analyzed.

Selection of test items

Both tests, Form A and Form B, consisted of words and non-words. Some of the words were from Stanford Achievement Test. Non-words were created by adding consonant letter(s), adding improper endings, and changing a vowel or consonant letter or letters. Out of 163 items in each form, 60% were words and 40% were non-words.

Procedure

Testing was carried out in groups of 25 students in their classroom. For each test form, subjects were given a list of 163 words and non-words which appeared on four sheets of paper. Some of the students took Form B first and then Form A. They were told that many of the words in the test were not really English words at all and were asked to respond "yes" to a word if they knew its meaning. They were asked to mark "T" or "F" on the computerized answer sheet.

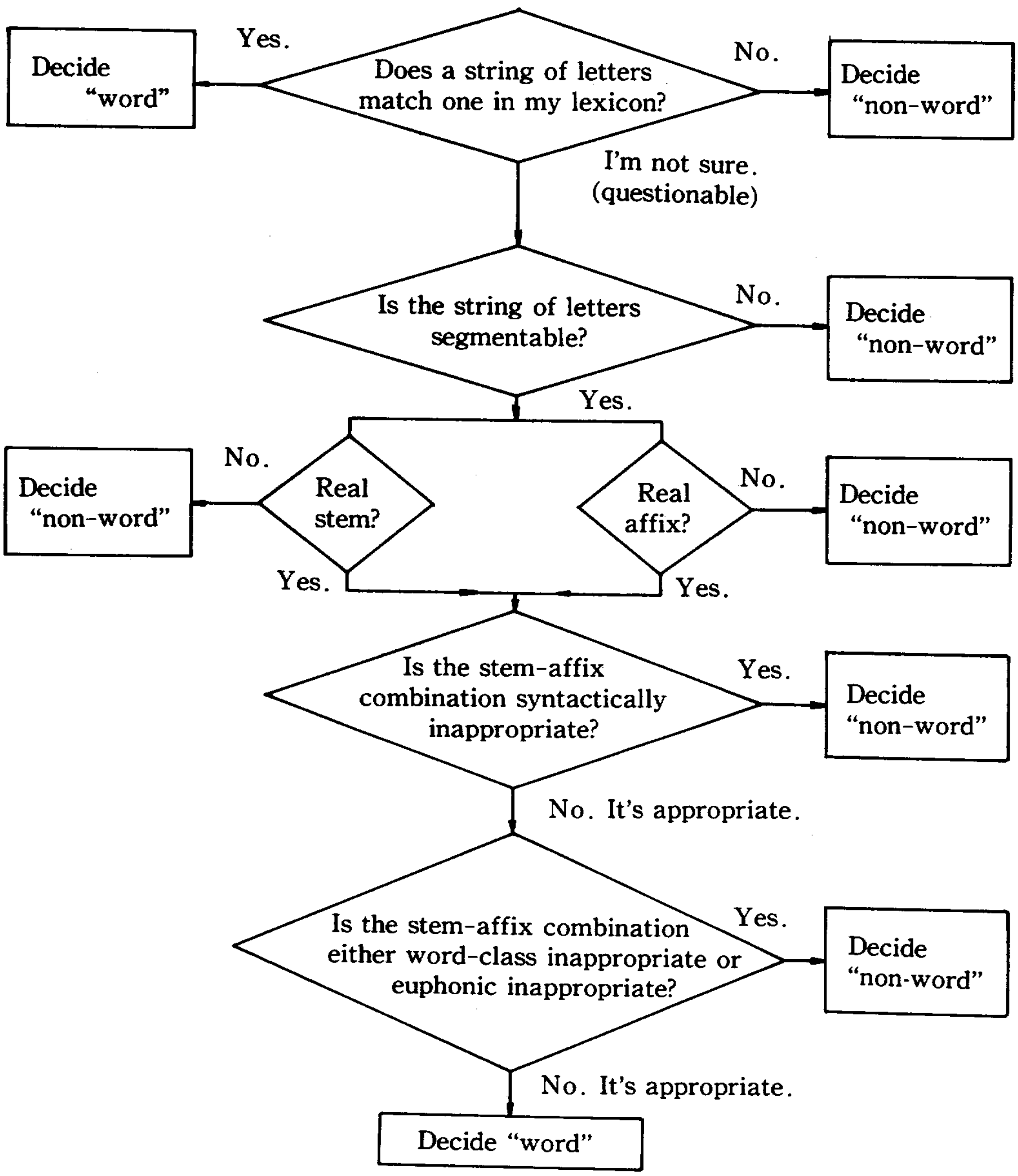
Coding system

For the purpose of this research, it is necessary to develop a system for coding words and non-words to detect the patterns and strategies in children's vocabulary acquisition. The coding system developed comprises the following four elements: (1) word/non-word, (2) segmentability, (3) error causing possibility for non-words, and (4) frequency for words. This coding system is based on the hypothetical cognitive operations to decide whether a given string of letters is a known word or a non-word. The hypothetical flow chart is as in Figure 1.

If we are given a string of letters as stimulus and are asked if we know its meaning or not, what happens in our mind? First, we ask ourselves whether it matches one in our lexicon. If it does we say, "Yes, I know the meaning", and if it does not, we say, "No". But in case we are not sure, we do further analyses. Some words are prefixed and/or suffixed, and if we can tell the meaning (or the function) of the prefix and/or the suffix and also if the stem (or base) after the affixes are removed matches one in our lexicon, we could say "Yes, I know the meaning of this word", even if we have not heard the word before. On the other hand, we are obliged to say "I don't know the meaning" if a word can't be separated into a stem and an affix and we can't find it in our lexicon. Thus, if the prefix and/or suffix are meaningful and the stem is real, but if the combination of these should not occur, this kind of "word" should not match one in our lexicon. Two stages are hypothesized for this decision. The first one is the case when the combination is syntactically inappropriate: for example, the suffix"-ly" does not go with a verb. The second is the combination which is syntactically possible, but is either word-class inappropriate or euphonic inappropriate. An example for this case is that the suffix "-tion" goes only with a Latin stem, not with others.

Figure 1

Hypothetical flow chart.
Aim: to decide if a string of letters is a word or non—word
〔meaningful or not〕



The definition of "segmentable" should be clarified here. Only when the word is separated into a free base — which is meaningful by itself — and a prefix/suffix part, or when it is a compound word, is it called a segmentable word. In other words, if a word does not have an immediate ancestor, it is not segmentable. An immediate ancestor of a word is defined as "the most closely related word which can in any sense be considered more 'basic' than the target word" (Anderson and Nagy, 1981). Therefore it should be pointed out that "segmentability" in this coding system is quite different from that in the customary linguistic or "linguistical" sense. The reason linguistical segmentability was not used for the coding is that the subjects were fifth graders and they were not assumed to perceive words in the way they are made up of bound base(s) (i.e., a Greek or Latin base) and a prefix and/or suffix. Since there is no principled way to assign the meanings of bound base and affixes, this linguistical approach which is more etymological or historical than synchronic is not appropriate especially for these subjects. The rule-of-thumb in breaking down the string is to remove a suffix rather than a prefix. For non-words, a possible existing word is thought about from a stimulus and its segmentability is determined according to that of the existing word. After an affix is removed, the stem which carries the principal meaning of a word and which does not have its immediate ancestor remains. Then, whether an affix and the stem are real or not should be coded.

There are several possible reasons why a person might say, "I know the meaning of this word", for a non-word. The coding for "Error-Causing Possibility" was developed to identify these. The basic assumption is that a person thinks about a real word which he/she has seen or heard before when looking at the given non-word. He/she might think it is a word he/she knows because it sounds like a word whose meaning he/she knows when he/she pronounces it silently. It might be a visually careless mistake, too. Since the student's response is simply "Yes" or "No", we can't tell exactly which word he/she thought about with a given non-word. Therefore, the coding for Error-Causing Possibility was based on all the possible words guessed by a few people. The possibilities of errors to get to the guessed word from the non-word were coded for each item. The position of this process within the cognitive operation tree (in Figure l) is right before the decision.

For words, their frequency is also coded; its value is the "F" appearing in the table of "The American Heritage Word Frequency Book" (Carroll, Davies, and Richman, 1971). "F" is the sum of all the frequencies of a word type appearing in the textbooks for grades 3 through 9. It is expressed as an integer.

The summary of the coding system mentioned above is in Table 1.

Table 1

System for coding words and non-words

1. word or non-word

   1..word   0..non-word

2. segmentable

   1..segmentable   0..non-segmentable

3. real stem

   1..real stem   0..non-real stem

4. real affix

   1..real affix   0..non-real affix

5. word-class inappropriate

   (for real stem plus real affix)

1..yes 0..no

6. syntactically inappropriate
(for real stem plus real affix)

1..yes 0..no

7. ECP-1: spelling error of weak vowel
(pronounced—ə—)

1..yes 0..no

8. ECP-2: confusion with articulation similarity (auditory), phonetinally similar: sharing some features and differing in minor features, e.g., d/t, k/g, v/f, č/ǰ.

1..yes 0..no

9. ECP-3: confusion with visual similarity, e.g., m/n, d/b/p/q, v/w, a/o.

1..yes 0..no

10. ECP-4v: for vowel, spelling error but same pronunciation

1..yes 0..no

11. ECP-4c: for consonant, spelling error but same pronunciation

1..yes 0..no

12. ECP-5: long ⟷ short vowel transformation

1..yes 0..no

13. ECP-6v: for vowel, spelling error and different pronunciation

1..yes 0..no

14. ECP-6c: for consonant, spelling error and different pronunciation

1..yes 0..no

15. ECP-7: analogy of irregular inflection

1..yes 0..no

16. ECP-10: back formation: bound stem is used as a free stem

1..yes 0..no

17. ECP-8v: omitting a pronounced vowel which should appear in a real word

    1..yes   0..no

18. ECP-8c: omitting a pronounced consonant which should appear in a real word

    1..yes   0..no

19. ECP-9: there is a vowel/consonant that should not be pronounced

    1..yes   0..no

20. frequency of a word

    frequency "F" for a word   0..non-word

Results

The scoring was done in two ways. One is for responding "meaningful/ non-meaningful," and the other is based on whether the response is correct or incorrect. The response was called correct if a subject answered that they knew the meaning to a word or they didn't know the meaning to a non-word.

Ability levels were determined by their total score. The high ability group is the top quarter, and low ability group is the bottom quarter of the total sample.

Multiple regression analysis

This analysis was done using 19 predictor variables (variable 2 through 20 in the coding) and the meaningful(=1)/ non-meaningful(=0) response as a criterion.

The squared multiple R and standardized regression weights were obtained for each subject based on the response in Form-A. The range of R^2 was .088 to .743 and the mean was .393. Also the correlation between R^2 and total score was .901.

The high ability students (top quarter) and the low ability students (bottom quarter) were compared using their mean weights. Significant differences were found in variable 2, 6 to 14, and 16 to 18, as shown in Table 2. This tells us that for good students the weight for real stem is very large in determining the meaningfulness of the stimulus while poor students proceed to error-causing possibilities more often except for ECP-10 (back formation).

Table 2

Comparison of means of multiple regression weights
between high and low ability group

	variable	high	low	$t(58)$	p
1.	segmentable	−.145	−.167	−.135	$p=.1684$
2.	real stem	.588	.274	13.592	$p<.00005$
3.	real affix	−.127	−.156	1.834	$p=.0718$
4.	word-class inappropriate	−.103	−.108	.312	$p=.7564$
5.	syntactically inappropriate	−.031	−.020	−.635	$p=.5281$
6.	(1)	−.299	−.187	−9.364	$p<.00005$
7.	(2)	−.287	−.096	−11.064	$p<.00005$
8.	(3)	−.344	−.115	−11.514	$p<.00005$
9.	(4 V)	−.376	−.202	−12.742	$p<.00005$
10.	(4 C)	−.242	−.097	−10.689	$p<.00005$
11.	(5)	−.220	−.055	−10.240	$p<.00005$
12.	(6 V)	−.463	−.243	−12.915	$p<.00005$
13.	(6 C)	−.370	−.130	−10.861	$p<.00005$
14.	(7)	−.235	−.127	−8.981	$p<.00005$
15.	(10)	−.094	−.055	−2.476	$p=.0191$
16.	(8 V)	−.075	−.042	−2.916	$p=.0085$
17.	(8 C)	−.185	−.070	−7.013	$p<.00005$
18.	(9)	−.167	−.094	−5.731	$p<.00005$
19.	frequency	.320	.315	.241	$p=.8121$

Error pattern analysis

Variables in the coding were used to categorize items into several groups. Based on the flow chart in Figure 1, the following four dichotomous variables were chosen for this purpose: (1) word/non-word, (2) segmentable/non-segmentable, (3) real stem/non-real stem, and (4) real affix/non-real affix. The reason for this choice is that the variables (2), (3), and (4) determine the major characteristics of the stimulus according to the hypothesis mentioned in this research and variable 1 was necessary to judge whether the response was correct or incorrect. The two variables about the characteristic of combination of a real stem and a real affix and also the variables for Error- Causing Possibility were not included for this because they seemed minor elements in categorizing items.

In the course of examining each student's response pattern by means of a response retrieval program, it was noticed that the profiles of Q values (the probability of an item being answered wrong) for the four variables were similar among high ability students. The four dichotomous variables mentioned above can make 16 patterns. In this analysis, however, only five patterns, each of which had more than 12 cases, were examined. Most of the items used in the tests were categorized into these five patterns. The Q-value profile for the high ability group was as shown in Table 3.

For each subject, the mean Q values for these five patterns were computed and, in addition, the extent to which an individual's response profile is typical or consistent with respect to the average pattern for the high ability group was measured. For measuring this, Tatsuoka and Linn's (1981) extended caution index was used. The extended caution index is a generalization of the caution index developed by Sato (1975).

Table 3

Average Q value profile for high ability group (criterion for Ci)

pattern		Q	
P 1	0111	.2944	(non-word, segmentable, real stem, real affix) e.g. adjustion, arousion
P 2	1111	.1954	(word, segmentable, real stem, no affix) e.g. creation, invisible
P 3	1010	.1258	(word, non-segmentable, real stem, no affix) e.g. forbid, anvil
P 4	0101	.0842	(non-word, segmentable, non-real stem, real affix) e.g. faminanity, ritter
P 5	0000	.0313	(non-word, non-segmentable, non-real stem, no affix) e.g. flane, cobe

A caution index is obtained from an S-P score table, which is a specially arranged binary data matrix. It is a measure of disparity between two sets: the observed one and the perfect one. Theoretically, it is expressed as the ratio of the disparity between actual and ideal response patterns and the maximum possible disparity from the ideal response pattern (Tatsuoka, 1978).

The extended caution index replaces the dichotomous scores by the Q-values, and thus, the formula used for the present analysis is as follows:

$$Ci = 1 - \frac{Cov\ (Qij\ ,\ \ Q.j)}{Cov\ (Qij^{*},\ \ Q.j)}$$

where Q.j is the descending ordered average Q values for high ability students, Qij is the actual Q values for the ith subject, and Qij^{*} is the Q values for the ith subject rearranged in descending order (Tatsuoka and Linn, 1981).

This index is a measure of the discrepancy of a subject's profile from the criterion profile. If a person's average Q-value profile is similar to the criterion one, Ci will be small. The more the person's response pattern deviates from this, the larger the Ci he/she will get. The correlation between Ci and total score is −.644. The higher Ci, the lower the total score in general.

A t-test for the mean Ci difference between high and low ability groups was also done. The high ability group (mean .2297, s.d. .2769) and low ability group (mean .8529, s.d. .3208) were compared and $t(58) = -8.045$ ($p < .00005$) was obtained. Thus, there was a significant difference of Ci between the two groups.

From the above-mentioned results, it can be concluded that the high ability students have similar profiles of response patterns based on the four dimensions, and the Q value profile of low ability students is significantly different from this. On the other hand, for low ability students, consistent patterns were not discovered using the variables of Error-Causing possibility. A coarser grouping in which, for example, variables 6, 9, 11, 12, and 16 (all involving vowel errors) occur, might lead to a consistent pattern.

Finally, the patterns of the most affirmed non-words for the high and low ability groups were examined. If Q was larger than .400 for the high ability group and larger than .140 for the low ability group (i.e., about 20 non-words for each group) a non-word was designated as "most affirmed." Table 4 separates out the non-words affirmed by both groups from those affirmed by one group only and gives the coding for each non-word.

In the hypothetical cognitive strategy, the ability to detect a syntactically inappropriate combination of a real stem and a real affix was supposed to be acquired earlier than that to detect the inappropriateness of a combination that is well-formed but never occurring, but this hypothesis was not supported in this study. $T(14) = -1.173$, $p = .2605$ for high abilhty group and $t(14) = -2.696$, $p = .0174$ for low ability group were obtained.

Table 4

Non-words most affirmed by (1) both groups; (2) the low ability group only; and (3) the high ability group only

loyalment	0111010000000000000	revese	0000000000001000010
successment	0111010000000000000	cobe	0000000010011100000
deformness	0111010000000000000	robbit	0000000010011000000
observement	0111100000000000000	weast	0000000000001000000
strangity	0111100000000000000	porfame	0000001001001000000
compure	0111010010001000000	grell	0000000001001000000
risent	0000001001100000000	sleem	0000000000010000000
jarbal	0000001001001100100	crang	0000000100000100000
slead	0000000100011100000	flane.	0000000010000100000
mudge	0000000010000100000	plode	0000000000010000000
		lote	0000000000010000000
		bighter	0101000000100000000
		ritter	0101000100110100000
		monder	0101000001001000000
		escore	0111010000000000010*
	(1)	(2)	

	(3)
conversal	0111100000000000000
adjustion	0111100000000000000
assistity	0111100000000000000
persistion	0111100000000000000
freshity	0111100000000000000
arousion	0111100000000000000
notator	0111100000000000000
instructness	0111010000000000000
forgivity	0111010000000000000
exhilirated	0101001000000000000
roversal	0101001001001000000

* Coding for "escore" is the mixture

of	0111010000000000000
&	0010000000000000010

Discussion

From multiple regression analysis results

The squared multiple R indicates "the proportion of variability in Y that is accounted for by the linear regression on the predictors in the normative sample" (Tatsuoka, 1969). When the meaningful/non-meaningful response was taken as the criterion, the squared multiple R was fairly high (range .088 to .743; mean .393), compared to the result when the criterion was correct/incorrect response, and the correlation of .901 between R^2 and total score shows that the higher the squared multiple R, the higher the total score. This implies that the coding system fits good students better than it does poor students. Good students could be following a flow chart similar to that hypothesized in Figure 1 when they decided if a string of letters was meaningful or not. Poor students' responses appear more random when the coding for the items is based on that hypothesis.

Generally, students' responses are determined by whether the stimulus has a real stem or not and also the extent to which they are familiar with the stimulus. But we notice that all the students' strategies to make a decision are not the same when we examine the regression weights for each subject. The asterisks in the regression equation indicate that the subject has a relatively larger weight for the component and help us interpret the general tendency of each student's strategy.

For example, the equations

$$Y = -.121X_1^{*} + .569X_2^{*} - .101X_3^{*} - .134X_4 + .111X_5 - .306X_6 - .252X_7 - .295X_8 - .370X_9 - .229X_{10} - .210X_{11} - .427X_{12}$$

$$-.334X_{13}-.201X_{14}-.103X_{15}-.069X_{16}-.159X_{17}-.156X_{18}$$
$$-.374X_{19}^{*} \quad \text{for subject \#147}$$

and

$$Y=-.211X_{1}+.405X_{2}-.199X_{3}-.133X_{4}-.086X_{5}-.217X_{6}^{*}$$
$$-.189X_{7}-.224X_{8}^{*}-.259X_{9}-.161X_{10}-.145X_{11}-.317X_{12}^{*}$$
$$-.239X_{13}^{*}-.153X_{14}^{*}-.068X_{15}-.046X_{16}^{*}-.120X_{17}^{*}-.100X_{18}^{*}$$
$$+.446X_{19}^{*} \quad \text{for subject \#8}$$

tell us that subject #147 is prone to decide that a presented word/non-word is meaningful when it is segmentable, has a real stem, has a real affix, or is a frequent word, while subject #8 tends to decide "Yes" when it has Error-Causing Possibilities 1, 3, 4C, 6V, 7, 8V, 8C, 9 or if it has a high frequency. In general, high ability students decide the meaningfulness of a stimulus by examining if it contains a real and meaningful word with or without a real affix. On the other hand, low ability students are easily misled by the spelling either in vowel or in consonant.

<u>From error analysis results</u>

In this analysis, the response "Yes, I know the meaning" to a non-word and the response "No, I don't know the meaning" to a word, were both counted as errors.

It was shown that high ability students' response patterns are more systematic than those of low ability students. The former group makes errors most often when the item is a non-word of a real-stem-plus-real-affix cmbination then for a word with an affix and then for a word without an affix. They don't make errors much when a non-word has a non-real stem. Low ability students' patterns deviate significantly from this and Q values for a word and a non-word without an affix discriminate high and low-ability groups very well.

It is true that the low ability students' profiles are significantly different from those of the high ability group. However, it is risky to conclude that students' Q-value profiles would be similar to the above mentioned result if they were given other forms of this kind of vocabulary test. The reason for this is that the frequency of words plays an important role when subjects make decisions on the meaningfulness of a word. The mean frequencies for 1111 type items and 1010 type items are 69.1 and 161.0, respectively. This frequency difference might have caused the order of difficulty of item types of words.

Table 4 shows how different the patterns of non-words affirmed by each group are. The types of non-words that attracted both groups are either visually or auditorily confusable ones. Their errors are interpreted with the vowel/consonant transformations used in Error-Causing Possibility, but they are not consistent. On the other hand, high ability students affirmed the non-words that consist mostly of a real stem and real affix. Their errors in visually or auditorily confused ones (e.g., "exhilirated" and "roversal") have a common error type. i.e., ECP−1. Their typical error occurs when a given non-word has a vowel spelling error in an unstressed vowel.

Implications

It became clear that the acquisition of vocabulary knowledge is not always the same as the acquisition of operation skills such as mathematical operation skills. Students do not always make errors (i.e., respond "yes" to a non-word or "no" to a word) when the items have the same possibility of leading to errors. But some generalizable rules were found from the results.

The strategy of vocabulary acquisition can be explained by the combination of two hypotheses. Up to a certain stage, invidual word meaning plays the main role in the acquisition. And gradually children acquire the conceptual framework concerning a root word and an affix. At the developmental stage of the subjects in this research, high ability students have acquired this conceptual framework better than low ability students. They are not easily confused by the visual or auditory similarity of pseudo-words to words. They judge the stimulus as a non-word if it does not possess a real stem and they have acquired the skill to guess the meaning from a root word and an affix.

Moreover, it was pointed out that everybody does not use the same decision rule. Their cognitive process cannot be interpreted easily.One reason for this is that if a person has never been exposed to a certain word before, there is no way to guess the meaning. The vocabulary acquisition situation is not the same for everybody.

The nature of an individual's error-making tendency, i.e., their cognitive process, was picked up, but the weaknesses of this research must be pointed out. First, as mentioned before, the yes/no response pattern to each item is not good enough to analyze errors. Information on what word or what meaning they think about is necessary for this purpose. Second, the items should be made and selected based on the hypothetical cognitive operations in order to test the hypothesis. It is desirable that each pattern on non-words have the same number of items. Third, to clarify the role of word frequency in vocabulary acquisition, the variable "frequency" used here in the coding system is not enough. Once the role of frequency is specified, the morphological acquisition and also the decision strategy will become more accurate. The major weakness of this research is that high and low ability groups were determined by the total test score of the vocabulary test itself. Thus, the concurrent validity of the test remained questionable.

The following points would be suggested for further research. First, the role of word frequency should be adjusted by use of covariance analysis. The frequency of a word with which a non-word is confused is probably positively correlated with the likelihood of affirming the latter as a word. After the adjustment of the role of word frequency, an ordered set of transformations will become less ambiguous. In other words, the hierarchy of the transformation that occurs, after each of which a match with a word in internal lexicon is sought, needs to be specified.

References

Anderson, R. C. and Freebody, P. *Vocabulary knowledge*. Technical Report No.136, University of Illinois, Urbana: Center for the Study of Reading, August, 1979.

Anderson, R. C. and Freebody, P. *On deciding if a string of letters is meaningful*. [Unpublished paper] University of Illinois, Urbana: Center for the Study of Reading, 1980.

Anderson, R. C. and Nagy, W. *System for coding relatedness among word type in the American Heritage word frequency book*. [Preliminary draft] University of Illinois, Urbana: Center for the Study of Reading, March, 1981.

Barganz, R. A. *The morphophonemic performance of good and poor readers*. Technical Report No.182, University of Wisconsin, Madison: R & D Center for Cognitive Learning, project in Elementary Reading, November, 1971.

Bergan, J. R., Towstopiat, O., Cancelli, A. and Karp, C. *Replacement and component rules in hierarchically ordered mathematics rule learning tasks*. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, April, 1981.

Brown, J. S. and Burton, R. R. Diagnostic models for procedural bugs in basic mathematics skills. *Cognitive Science*, 1978, 2, 155-192.

Carroll, J. B., Davis, P., and Richman, B. *The American Heritage word frequency book*. New York: American Heritage Publishing Co., 1971.

Fredericksen, J. R. *Assessment of perceptual, decoding, and lexical skills and their relation to reading proficiency*. Report No.3756, Technical Report No.1, Bolt Beranek and Newman Inc., January, 1978.

Johnson, D. D., Pittelman, S. D., Schweiker, J., Shuberg, L. K., and Morgan Janty, C. *Relationships between work identification skills and reading comprehension of elementary school children*. Technical Report No.490, University of Wisconsin, Madison: R & D Center for Individualized Schooling, December, 1978.

Langman, M. P. The Reading process: a descriptive, interdisciplinary approach. *Genetic Psychology Monographs*, 1960, 62, 3-40.

Rhode, M. and Cronnell, B. *Compilation of a communication skills lexicon coded with linguistic information*. Technical Report 58, SWRL Educational Research and Development, November, 1977.

Rips, L. J., Shoben, E. J., and Smith, E. E. Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 1-20.

Rubenstein, J., Lewis, S., and Rubenstein, M. A. Evidence for phonetic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior*, 1971, 6, 645-657.

Sato, T. *The construction and interpretation of S-P tables*. Tokyo: Meijitosho, 1975. [In Japanese]

Tatsuoka, K. K., Birenbaum, M., Tatsuoka, M. M., and Baillie, R. *Psychometric approach to error analysis on response patterns of achievement tests*. [Technical Report 80-3] Urbana, University of Illinois, Computer-based Education Research Laboratory, 1980.

Tatsuoka, K. K., and Linn, R. L. *Indices for detecting unusual response patterns: links between two general approaches and potential applications*. [Reserch Report 81-5], Urbana, University of Illinois, Computer-based Education Research Laboratory, 1981.

Tatsuoka, M. M. *Validation Studies: The use of multiple regression equations*. Champaign, Illinois: Institute for Personality and Ability Testing, 1969.

Tatsuoka, M. M. *Recent psychometric developments in Japan: Engineers grapple with educational measurement problems*. Paper presented at the ONR Contractors' Meeting on Individualized Measurement, Columbia, Missouri, September 19, 1978.

Walker, J. H. Pronounceability effects on word-nonword encoding in categorization and recognition tasks. *Journal of Experimental Psychology*, 1973, 99, 318-322.