Variability of Error Patterns on Two Formats of Picture-to-Word Matching

Craig W. Linebaugh
The George Washington University, Washington, D.C.

Clinical aphasiologists have in recent years become increasingly concerned with reading and the clinical management of alexia. Two widely-used tests which may provide useful information for focusing alexia treatment are the Minnesota Test for Differential Diagnosis of Aphasia (Schuell, 1965) and the Reading Comprehension Battery for Aphasia (RCBA) (LaPointe and Horner, 1979). Both of these tests contain subtests of picture-to-word matching in which the response choices for a given picture stimulus are related to the target response in a particular way; i.e., visually, auditorily or semantically. In both tests, only one foil type is presented for each stimulus, thus limiting the types of discriminations a subject need make on any given stimulus and increasing the probability of making an error of a particular type. Of both theoretical and clinical relevance is the following question. Would the information obtained from picture-to-word matching tasks be more valid if subjects needed to contend with visually, auditorily and semantically related foils on each stimulus? This study compared two picture-to-word matching formats, one in which all the foils for each stimulus were of one type and one in which each stimulus included three different types of foils.

METHOD

Subjects. The subjects for this study were 25 aphasic patients. Their mean age was 55.9 years with a range of 17-82 years. Their mean educational level was 11.5 years with a range of 4-19 years. Months post onset of aphasia ranged from 1-149 with a mean of 36.0 months. Mean aphasia quotient determined with the Western Aphasia Battery (WAB) was 64.4, with a range of 4.3-93.8. The subject population included 6 Broca, 2 Wernicke, 12 anomic, 2 conduction and 3 global aphasic patients, based on WAB classifications.

Procedure. The experimental task used in this study consisted of a picture-to-word matching task in which subjects selected which of four printed words best matched a black and white drawing. The pictures used were those of subtests I, II and III of the RCBA. The printed words were presented in 24 point helios bold lower-case type.

Two different forms of the picture-to-word matching task were used. In the One Foil Type (OFT) format, all three foil response choices were either visually, auditorily or semantically related to the target response. In the Three Foil Type (TFT) format, the foil responses included one which was visually similar to the target response, one which was auditorily similar and one which was semantically similar. In the OFT format, two of the foils were taken from the RCBA and a third foil was selected in accordance with the RCBA specifications for foil selection for subtests I, II and III. In the TFT format, one foil of the appropriate type was selected from the RCBA and two additional foils were selected in accordance with the RCBA specifications.

A different randomization of the 30 stimuli was developed for each format. Each format was administered twice in counterbalanced alternating order to the 25 subjects. Administrations of the task were separated by at

least four days, and no more than three weeks elapsed over the four administrations of the task.

RESULTS

Of the 25 subjects tested, seven made 2 or fewer errors on each administration of the experimental task. These subjects were not included in the analysis of error patterns. In addition, another subject who was one month post onset made a rapid recovery in all modalities during the testing period, and, therefore, his results were also excluded from further analysis. Thus, data from 17 subjects were used for the analysis of error patterns.

Group Error Patterns

Table 1 shows the mean number of total errors and the mean percentage of errors of each type; i.e., visual, auditory or semantic, on the first and second administrations of the OFT and TFT formats. Percentages of error types were used because the number of opportunities to make an error of a particular type differed between the two formats.

Table 1. Mean (standard deviation) number of total errors and mean percentage of errors of each type for the first and second administrations of each format.

	One Foil Type		Three Foil Type	
	1st administration	2nd administration	1st administration	2nd administrat
Total	7.1 (5.54)	6.2 (5.16)	7.8 (5.29)	6.9 (4.69)
Visual	48.5 (23.08)	47.3 (23.55)	44.1 (22.97)	48.7 (27.83)
Auditory	20.7 (14.87)	24.7 (17.48)	11.6 (15.17)	13.5 (14.61)
Semantic	30.8 (17.60)	27.9 (26.69)	44.3 (23.99)	37.8 (27.61)

For statistical analysis, the means shown in Table 1 were organized into three overlapping clusters. Two-tailed, matched-pair t tests for related samples were used to compare certain means in each cluster. To maintain a reasonable error rate across these multiple comparisons, an alpha level of .05 divided by the number of comparisons within a cluster was used (Dunn, 1961).

Comparisons between Formats. The means for total number of errors and percentage of errors of each type were compared between the OFT and TFT formats. This involved four comparisons with an alpha level of .0125. None of the differences between the two formats was statistically significant.

Comparisons among Error Types in Each Format. The mean percentages of errors of each type were compared within the first administration of each format. This involved three comparisons within each format, for a total of six comparisons, with an alpha level of .008. On the OFT format, the subjects made significantly more visual than auditory errors ($t=3.31$, $p<.004$). The differences between the mean percentages of visual and semantic errors

and between those of auditory and semantic errors were not statistically significant. On the TFT format, the subjects made significantly more visual than auditory errors ($t=4.37$, $p<.001$) and significantly more semantic than auditory errors ($t=4.09$, $p<.001$). The difference between the percentages of visual and semantic errors was not statistically significant.

Comparisons between First and Second Administrations of Each Format. The mean total number of errors and the mean percentages of errors of each type were compared between the first and second administrations of each format. This involved four comparisons within each format, for a total of eight comparisons with an alpha level of .006. None of the differences between any of the means compared was statistically significant.

Correlations. Pearson product-moment correlation coefficients were computed for the total number of errors and the percentages of errors of each type between the two formats. These coefficients are shown in Table 2. The coefficient for total errors was a relatively high .80 ($p < .001$). Those for the error types, however, were all low and nonsignificant.

Table 2. Pearson product-moment correlation coefficients for mean total number of errors and mean percentages of errors of each type between the two test formats.

	One Foil Type			
	Total	Visual	Auditory	Semantic
Three Foil Type				
Total	.80			
Visual		.33		
Auditory			.08	
Semantic				-.02

Correlation coefficients were also computed for total number of errors and percentages of errors of each type between the first and second administrations of each format. These coefficients are shown in Table 3. Relatively high correlations were obtained for total errors for both the OFT ($r=.79$, $p<.001$) and TFT ($r=.89$, $p<.001$) formats. The correlations for error types in both formats were all low and nonsignificant.

Table 3. Pearson product-moment correlation coefficients for mean total number of errors and mean percentages of errors of each type between first and second administrations of each test format.

	Total	Visual	Auditory	Semantic
One Foil Type				
Total	.79			
Visual		.44		
Auditory			.40	
Semantic				.25
Three Foil Type				
Total	.89			
Visual		.39		
Auditory			.36	
Semantic				.30

Individual Error Patterns

Ten subjects made more errors on the TFT format than on the OFT format. Two subjects made equal numbers of errors on both formats, and five made more errors on the OFT format. Only five subjects showed the same relative percentages of errors of each type on the two formats.

On both formats, seven subjects made greater than 50% of their errors of one type. On the OFT format, six made greater than 50% visual errors and one made greater than 50% semantic errors. On the TFT format, two made greater than 50% visual errors and five made greater than 50% semantic errors. Only two subjects made greater than 50% of their errors of the same type on both formats.

Comparing the two administrations of the OFT format, 11 subjects made fewer errors, two made an equal number of errors and four made more errors on the second than the first administration. Only eight subjects showed the same relative percentages of errors of each type on the two administrations. On the two administrations of the TFT format, nine subjects made fewer errors, two made an equal number of errors and six made more errors on the second than the first administration. Only six subjects showed the same relative percentages of errors of each type on the two administrations.

DISCUSSION

Overall, the subjects in this study made slightly more visual than semantic errors and significantly more visual and semantic errors than auditory errors. These results are generally in agreement with previous studies regarding the distribution of errors among error types. Gardner and Zurif (1976) reported that their 36 aphasic subjects as a group made more errors on a task requiring selection of one of four semantically-related words to match a picture than on a task requiring selection of one of four auditorily-related words. The difference between the mean numbers of errors on these two tasks was not significant, however. Webb and Love (1983) used a task in

which subjects selected one of six words to match a picture. The six words in the response field included visual, auditory and semantic foils. Their subjects made semantic errors most frequently, followed by visual errors and then auditory errors, but no statistical analysis to determine the significance of the differences among the error types was reported. Van Demark, Lemmer and Drake 1982) examined the performance of 26 aphasic patients on the RCBA. They found that their subjects made significantly more errors on subtest 1, visual foils, and subtest 3, semantic foils, than on subtest 2, auditory foils. The numbers of errors on subtests 1 and 3 did not differ significantly.

Of greater interest in the current study, however, is the variability among the distributions of individual subjects' errors between the two formats. From a clinical and research standpoint, the most important findings were (1) on both formats only 7 of the 17 subjects included in the error analysis made more than 50% of their errors of a given type; (2) on the OFT format, six subjects made more than 50% visual errors and one made more than 50% semantic errors; (3) on the TFT format, two subjects made more than 50% visual errors and five made more than 50% semantic errors; (4) two subjects made more than 50% visual errors on both formats; and (5) two subjects made more than 50% visual errors on the OFT format and more than 50% semantic errors on the TFT format. The overall trend was for more subjects to make a majority of visual errors on the OFT format and a majority of semantic errors on the TFT format. Therefore, the questions that need to be addressed are (1) What are the sources of this variability? and (2) Which of the two formats provides a more valid assessment of patients' single-word recognition?

In considering possible sources of variability, one must first acknowledge the inherent variability of the language performance of aphasic individuals. Seven subjects included in this study failed to make more than 50% errors of any given type on either format. This group included the three subjects who made the greatest number of errors on both formats, but it also included three subjects who made six or fewer errors on both formats. As Kertesz (1979) has observed, "In most cases of aphasic alexia, the error types are multiple and the distinction is blurred as to which strategy of reading is impaired. The semantic, phonological and orthographic factors of reading are often affected together..." (p.210).

Kertesz' observation is not applicable to all of our subjects, however. Ten subjects did demonstrate some consistency in their performance on at least one of the formats. This group included five of the eight subjects who made six or fewer errors on both formats, but it also included five subjects who made at least seven errors on one of the formats. Perhaps these subjects used some task strategy which promoted a particular type of error on at least one of the formats.

There are two main strategies subjects might employ in performing a picture-to-word matching task. Both of these strategies lend themselves to either a direct selection approach, in which response choices are not compared with one another, or an elimination approach, in which response choices are compared with one another.

The first of these two strategies, which might be referred to as a "recall" strategy, is shown in Figure 1. A subject using this strategy first attempts to label the pictured object. Having retrieved an appropriate word, or perhaps an inappropriate word, he then either directly accesses a visuographic word image or constructs a visuographic image by grapheme selection and sequencing or by applying phoneme-to-grapheme conversion rules. The

visuographic image can then be compared to each of the response choices and the best match selected as the response. Some subjects may elect to compare pairs of response choices to the visuographic image, successively eliminating one member of each pair until the target response is isolated. In this strategy, auditory errors might occur if there is some breakdown in phonological encoding. Semantic errors might occur if a semantically-related word is retrieved rather than the target word, particularly if the semantically-related word is included in the response field. Visual errors could occur with breakdowns in grapheme selection and sequencing, in phoneme-to-grapheme conversion or when comparing the visuographic image to individual response choices or to pairs of response choices.

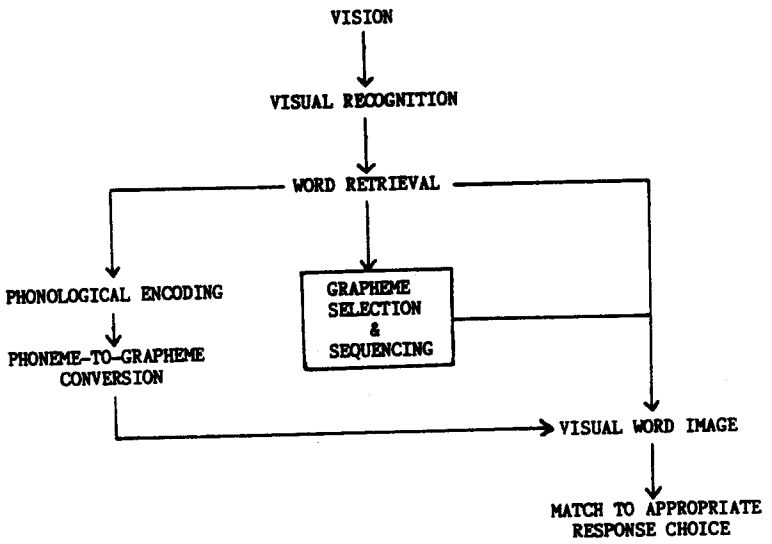

Figure 1. Recall strategy for performing the picture-to-word matching task.

Figure 2 presents the second main strategy which subjects might employ. It can be referred to as a "recognition" strategy. Here the pictured object is recognized, thus evoking a set of semantic features and associations. The subject then considers the response choices, selecting the one which evokes a similar set of semantic features. Current neurolinguistic models of reading postulate that there are three primary routes by which words are recognized. One is a direct access from the printed word to the semantic field. A second passes through a lexical processor in which the visuographic image is recognized, which in turn permits access to the semantic field. A third involves grapheme-to-phoneme conversion with processing then proceeding through auditory recognition channels. In this model, auditory errors might occur with breakdowns in grapheme-to-phoneme conversion or auditory recognition. Semantic errors could occur when a semantic foil evokes a sufficient number of the semantic features of the pictured object to induce selection of the semantic foil. Visual errors seem less likely with this strategy, but could occur if one of the associations evoked by the picture is a visuographic image of a word to which a visual foil is then incorrectly matched. As with the recall strategy, subjects might elect to compare pairs of response choices with the

pictured object, again eliminating one member of each pair until the target is isolated.

Each of these strategies would seem to promote errors of a particular type. The recall strategy would seem to lend itself more to visual errors. Subjects could experience breakdowns in either the retrieval or formulation of a visuographic image or in matching this image to the response choices. Semantic and auditory errors would most likely occur only if a breakdown in word retrieval or phonological encoding led to formulation of a visuographic image which corresponded to a particular foil. The recognition strategy would seem to promote semantic errors. Here both target responses and semantic foils would evoke overlapping semantic features common to the pictured object. If the semantic features which were evoked did not sufficiently differentiate the target from the semantic foil, the subject might well select the semantic foil.

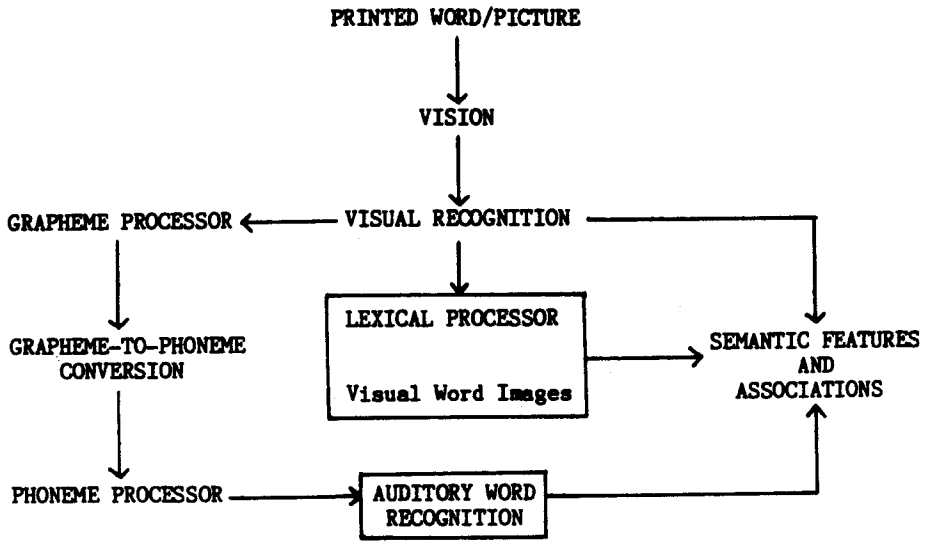

Figure 2. Recognition strategy for performing the picture-to-word matching task.

The question then arises as to why subjects might have employed a recall strategy which promoted visual errors on the OFT format and a recognition strategy which promoted semantic errors on the TFT format. An answer to this question may lie in the composition of the response field. In the TFT format, the foils were distinct from one another. Pairs of response choices requiring discrimination; i.e., the target and one foil, were apparent, and each discrimination could be performed on a distinctive basis--visual, auditory, or semantic. In the OFT format, however, the foils were related to one another, as well as the target, in a common way. The target-foil pairs to be discriminated were not so readily apparent, and as many as six discriminations of a given type might need to be executed. Discriminations among four similar response choices might be simplified by comparing a visuographic word image with each response choice, thus leading some subjects to use a recall strategy. On the other hand, the discriminations required on

the TFT format might be more readily accomplished than those on the OFT format using a recognition strategy. Regardless of the strategy employed, however, consistent performance, whether accurate or inaccurate, depends on the consistency with which the processes subserving the strategy are executed.

Another possible source of error pattern variability may lie in the nature of the associations evoked by particular foil types. Semantic foils evoke only semantic features and associations common to the target response. Visual foils, on the other hand, evoke not only common visuographic elements, but in the case of visual foils which share common first or first two or three graphemes with the target, they might also evoke an auditory association. Certain auditory foils also share common graphemes with the target word, but these graphemes tend to be at the end of the word and thus may be less salient in the recognition process. This combined visual-auditory effect might be particularly powerful when one is confronted with three visual foils, thus increasing the relative incidence of visual errors on the OFT format.

The results of this study provide no definitive basis on which to assert the greater clinical validity of one format over the other. Neither format has clearly superior test-retest reliability. The TFT format does provide a greater number of opportunities for each type of error to be made, but this is useful only if the data obtained provide a valid picture of a patient's processing strengths and weaknesses. If, as suggested above, the juxtaposition of three visual foils on certain stimuli in the OFT format promotes visual errors over auditory or semantic errors, the OFT format may yield a skewed picture of patients' word recognition abilities. The TFT format, therefore, may provide a more valid basis for analyzing a patient's deficits and for focusing intervention. Additional support for the TFT format may be garnered if, in fact, the difference in error distribution between the two formats is indicative of subjects' use of recognition strategies in the TFT format and recall strategies in the OFT format. Should this prove to be the case, the TFT format more closely approximates most functional reading tasks, wherein meaning is extracted from the printed word. The OFT format may, however, provide potentially useful information about patients' abilities to perform certain specialized reading tasks, such as using a dictionary or telephone book.

ACKNOWLEDGMENTS

The author wishes to acknowledge Beth Hillis Trupe for her assistance in data collection and Penelope Starratt Myers, Willie P. Cupples and Cheryl Person Margulies for their contributions to the discussion.

REFERENCES

Dunn, O.J. Multiple comparisons among means. Journal of the American Statistical Association, 56, 62-64, 1961.

Gardner, H. and Zurif, E. Critical reading of words and phrases in aphasia. Brain and Language, 3, 173-190, 1976.

Kertesz, A. Aphasia and Associated Disorders: Taxonomy, Localization, and Recovery. New York: Grune and Stratton, 1979.

LaPointe, L.L. and Horner, J. Reading Comprehension Battery for Aphasia. Tigard, OR: C.C. Publications, 1979.

Schuell, H. Minnesota Test for Differential Diagnosis of Aphasia Minneapolis, MN: University of Minnesota Press, 1965.

Van Demark, A.A., Lemmer, E.C.J. and Drake, M.L. Measurement of reading comprehension in aphasia with the RCBA. Journal of Speech and Hearing Disorders, 47, 288-291, 1982.

Webb, W.G. and Love, R.J. Reading problems in chronic aphasia. Journal of Speech and Hearing Disorders, 48, 164-171, 1983.

DISCUSSION

Q: What do you make of the low correlations you reported? Do they have some implications for the kind of dysfunction that these patients have? What I'm thinking about is the fact that a patient may make one kind of error on a word and then a different kind of error on the same word.

A: I think what it speaks to most directly is the inherent variability in the performance of brain-damaged individuals. I don't think any of us would suggest that a process is blown out entirely, except perhaps in the most severely involved patients, so that you would see the same kind of error again and again. In fact, most patients respond differently on different trials and that is what the low correlations are indicative of.

Q: Have you looked at other performance in your patients which could help you reduce the variability?

A: That's where we're going next. The purpose of the study was to compare two specific testing formats. This is one task one might use to focus intervention for alexia. The purpose was to try to determine which of the two formats is more reliable and more valid. Obviously, in focusing treatment for alexic patients, we want to look at a variety of tasks. There are many other tasks which could be done. We've only begun to scratch the surface with these subjects.

Q: In the RCBA, the semantic errors are in a way the least in error, the least obviously incorrect. When you have a picture of a basket, and you have basket and straw and eggs [as response choices], when one chooses straw, that's not as incorrect as if one chooses an auditory or visual foil. My guess is that some of those patients made more semantic errors in the TFT format because they were nearly correct responses. Straw is sort of correct for a straw basket. Yet when they saw four semantically-associated words, they said they're all correct, and I have to narrow it down to find the most correct. So that's just another possibility.