

Word Fluency and Aphasia: Some Linguistic
and Not-So-Linguistic Considerations

Michael Collins, Malcolm R. McNeil, Susan Lentz
Yvonne Shubitowski and John C. Rosenbek
Wm. S. Middleton Veterans Administration Hospital, Madison, Wisconsin

In 1967 Borkowski, Benton and Spreen asked three groups of subjects (including 30 brain damaged subjects) to produce as many words as possible, in one minute, beginning with each of 24 different letters of the alphabet. In this study, they attempted to determine the relationship between the fluency of controlled association to single letters and the frequency of words beginning with those letters in the English language. They also examined the extent to which verbal production prompted by association to specific letters differentiated between brain-damaged subjects and their normal controls.

Borkowski, Benton and Spreen (1967) found that the number of words produced correlated highly and positively with a frequency count derived from the Thorndike-Lorge norms and with estimates of the number of words in the English language beginning with each letter. They also found that this measure was particularly effective in differentiating brain-damaged from non-brain-damaged subjects.

That study generated substantial interest in word fluency for the differential diagnosis of brain-damaged from non-brain-damaged patients. Wertz and Lemme (1974) for example, collected normative data on 131 left hemisphere damaged patients, 40 right hemisphere damaged patients, and 63 normals. Although the results of that study have not been published, their normative data is widely available, and the results have been reported in several more recent studies (e.g. Wertz, Collins, *et al.*, 1981). Nevertheless, the value of the test may be more in its role as part of a pattern of response to a variety of speech and language tasks than as an independent, discriminating measure.

A particularly enduring belief in aphasiology is that all aphasic patients suffer some reduction in their available lexical repertoire, and that the words they do have available are those which are most frequent in the language (Goodglass and Kaplan, 1983; Howes, 1964). There is considerable evidence to suggest that this is true. Naming tasks, for example, typically yield better performance on "more frequent" items such as those in the Boston Naming Test (Kaplan, Goodglass and Weintraub, 1983) and, as patients recover, their performance on the more difficult items improves (Wertz, 1979). The Word Fluency Measure correlates highly with other language measures such as the Porch Index of Communicative Ability (Porch, 1967) and the Token Test (DeRenzi and Vignolo, 1962; Wertz, Collins, *et al.*, 1981), and thus possesses some concurrent validity as a measure for one aspect of aphasia.

Our interest in the Word Fluency Measure's potential for differential diagnosis, our uncertainty about what to make of the data once collected, and Darley's persistent question, "But what does it measure, Terry?" led us to the present study.

Two general queries governed the experimental questions we asked. The first was whether a more careful examination of the words produced would result in a more sensitive test to differentiate between brain-damaged and

normal groups. The second was whether a more detailed linguistic examination of these data would allow the differentiation of subtypes of aphasia.

We asked the following questions:

1. Do Broca's, Wernicke's, and mixed or unclassified aphasic patients and normal subjects differ in total number of words produced?
2. Does the average relative frequency of words produced differ among aphasic types or from normal productions?
3. Do average relative frequency scores change from initial to final test for aphasic subjects?
4. Do aphasic subjects differ from normal subjects in the production of words by semantic category chain, initial and final blends, or part of speech?

METHOD

Four subject groups were selected for this investigation: a normal, non-brain damaged group, a Broca's aphasic group, a Wernicke's aphasic group, and an unclassified aphasic group. All subjects in the aphasic groups were tested with the Word Fluency Measure at approximately one month post onset (the initial test), and at greater than six months post onset, plus or minus about two weeks (the final test).

There were 29 normal control subjects. Twenty-one aphasic subjects were unclassified, eleven were classified as Wernicke's aphasic, and thirteen were classified as Broca's aphasic. The criteria for classification were those of the Boston Diagnostic Aphasia Examination Rating of Speech Characteristics. Aphasic subjects ranged in age from 29 to 80, with a mean of 54.20 years. Normal subjects ranged in age from 26 to 80 with a mean of 51.26 years.

In administering the Word Fluency Measure all subjects were instructed to provide as many words as possible beginning with the letters "s," "t," "p," and "c," in one minute, and to exclude all proper nouns. The examiner then recorded and transcribed all words produced. Data were categorized among the following parameters. (1) Total number of words produced for the letter "s" only. The letter "s" was chosen because it was found to be the easiest for most aphasic subjects and we chose to optimize the number of responses for subsequent linguistic analyses. (2) The total number of semantic category chains. An example of a semantic category chain would be "soup, salad, sandwich" for the supraordinate of "food." (3) The number of initial blend sound-alike and final blend/sound-alike chains -- for example "ship, shape, shoot" for initial blends or "sand, send, spend," for the final blends. Final consonant repetitions such as "sip" or "sap" were also included in this category. We also measured the number of words by part of speech; that is, by noun, verb, adjective, preposition, etc. (4) The average relative frequency of use for all words combined, using Jones and Wepman's (1965) spoken word count, list B. The mean frequency of occurrence was obtained for each word and those figures were combined. This sum was then divided by the number of words produced, to yield the mean relative word frequency for each subject. For example, one subject produced 11 words. Seven of these words did not have a frequency value and therefore earned a value of 0. One of the remaining four words had a frequency value of 9.74, one had a frequency value of .75, another .44, and the final one had a value of 2.87. These figures were summed and the sum divided by the total number of words produced, yielding a relative average use score of 1.24 for that subject.

RESULTS

Average Number of Words Produced. No significant differences were found among the three aphasic groups in the average number of words produced for the letter "s." Total production was slightly greater for unclassified subjects (8.95) than for Wernicke's (7.64) or for Broca's (6.69) subjects. Predictably, normal subjects produced a significantly greater number of words ($p < .01$) than any aphasic group.

Average Relative Use. Among the aphasic groups, Broca's subjects had the highest score for average relative use (2.49), followed by the unclassified (.916), and Wernicke's (.248) groups. Normal subjects produced words with the lowest average relative use (.103). Average relative use was significantly lower for normals when compared to the Broca's and unclassified subjects.

Average relative use was then compared from initial to final test. Seventy-three percent of the unclassified and 71 percent of the Broca's aphasic subjects produced a relative use score which decreased (that is, words produced at final test were "less frequent" than those produced at initial test). None of our Wernicke's aphasic patients showed a similar tendency.

We also examined relationships among aphasia type and average relative use. First, we compared mean word use divided by number of words produced, for each group at initial and final test. We found a slight tendency for word frequency to decrease from first to last test (2.49 to .64 for Broca's and .92 to .12 for unclassified subjects) but that relationship was not consistent. The unclassified and the Broca's subjects produced significantly more frequently occurring words. The Wernicke's subjects produced significantly less frequently occurring words. The other groups were not significantly different from one another. Relative use increased for Wernicke's subjects (.25 to .47). None of the changes were significant ($p > .05$). Performance on the PICA improved for all subjects from initial to final test. Average relative use scores for each group at initial and final test are shown in Table 1.

Table 1. Average relative use for normal and aphasic subjects.

Group	Initial Test	Final Test
Unclassified	.916	.115
Wernicke's	.248	.472
Broca's	2.490	.640
Normals	.103	-

Finally, we computed Pearson product-moment correlations among groups for PICA overall scores and average relative frequency of use for each subject. All correlations were low and none were significantly different from a zero correlation ($p > .05$).

Next we analyzed relationships among semantic category chains, initial and final blend/sound-alike chains, and part of speech for all groups. These data were not analyzed with inferential statistics because of some concern for their use with normalized data, and because variability within the aphasic groups probably prevents these average differences from being reliable ones.

Semantic Category Chains. Our unclassified aphasic subjects were more likely to use semantic category chains than the other two aphasic groups. Normal subjects produced substantially more semantic chains than the aphasic groups. The normal subjects consistently produced semantic chains, and these differences are meaningful and probably reliable.

Initial Blend Sound-Alike Chains. Unclassified aphasic subjects employed an initial blend chain 18 percent of the time. Wernicke's subjects never produced such a chain, and Broca's subjects produced 29 percent of their responses in association with such a chain. Normal subjects were much more likely to produce strings of words beginning with the same blend. Following at least six months of physiological recovery and treatment, aphasic subjects produced about as many initial blend chains as normals did. Mean differences among the aphasic groups were small and individuals within groups overlapped.

Next, we compared initial blend sound-alike chains from initial test to final test. They reflected a tendency for more frequent use of this strategy as language ability improved and as patients were further from onset. Use of this strategy increased 38 percent for unclassified aphasic subjects, 60 percent for Wernicke's, and 28 percent for Broca's aphasic subjects. These results are shown in Table 2.

Table 2. Initial blend sound-alike chains.

Group	Initial Test (%)	Final Test (%)
Unclassified	18	56
Wernicke's	0	60
Broca's	29	57
Normals	72	-

Final Blend Sound-Alike Chains. Final blend sound-alike chains were relatively infrequent, averaging about 20 percent for all groups of subjects. They were essentially unchanged at six months post onset, except that the Broca's group decreased to zero their production of these chains. Normal subjects performed similarly to the aphasic groups in number of final blend chains produced.

Parts of Speech. In Table 3, it can be seen that nouns were produced more frequently and to a similar degree by all four groups. A similar pattern was followed by all groups for verbs and adjectives. Other parts of speech were produced, but they did not constitute a substantial proportion of the total. Again, the average percentage for each part of speech was similar among all four groups. However, substantial differences existed among the number of nouns, verbs and adjectives produced.

Table 3. Number of words by part of speech for normal and aphasic subjects.

Group	Nouns	Verbs	Adjectives
Unclassified	55%	34%	9%
Wernicke's	48%	30%	12%
Broca's	52%	29%	13%
Normals	54%	29%	14%

When we compared production of simple and compound words, we found that 98 percent of the words produced by our normal subjects were simple words. Unclassified aphasic and Broca's aphasic subjects showed a similar preference for simple words, but Wernicke's aphasic subjects reversed this trend, producing 29 percent simple words at initial test and 71 percent at final test. These results are shown in Table 4.

Table 4. Percent simple words produced for normal and aphasic subjects.

Group	Initial Test	Final Test
Unclassified	100	99
Wernicke's	29	71
Broca's	80	90
Normals	98	-

SUMMARY

Our results suggest that: (1) There were no significant differences among aphasic groups in total number of words produced for the letter "s." (2) None of our three groups of aphasic subjects employed a consistent strategy initially (e.g. semantic word chain, initial sound/sound-alike chain) but there was a tendency for such strategies to increase as general language ability improved. (3) Groups did not differ in their production of words by part of speech. (4) No significant differences were obtained among our aphasic groups for word frequency scores from initial to final test, but there was a tendency, at least in the Broca's and unclassified aphasic groups, to produce "more frequent" words initially and "less frequent" words as general language ability improved. The Wernicke's group produced more frequently occurring words as language ability improved. The reasons for this are not clear. One possibility is that more useful words occur more frequently, and initially the Broca's patients had more useful words at their disposal. This is a fragile theory at best, and deserves closer attention. (5) No significant relationships were detected between aphasia type and relative word use.

Our results did not yield the discriminating data we had predicted. It may be that the letter "s" (the "easiest" letter of those that Borkowski, Benton and Spreen examined) does not stress the system enough to permit greater differentiation among aphasia subtypes. More "difficult" letters, such as "Q" or "J" might permit clearer differentiations among subtypes of aphasia. Perhaps these data reflect a fundamental invalidity in the classification system that we used, or perhaps the classification system is simply unrelated to this type of lexical access.

We tried to project some implications for therapy from these data. We failed. Simply illustrating that aphasic and normal subjects utilize semantic chains or initial or final blend chains does not provide much impetus to say that they would have done better or worse if they had not used them. Further, we do not know that aphasic people can learn to use these chains.

We failed to differentiate groups of aphasic subjects by any of our analyses. No one subject within any aphasic group always or never produced a particular response type -- group and individual performance failed to

differentiate among groups. We need to collect more data, and we need to refine our patient samples. When we do so, perhaps our data will be useful as well as interesting.

REFERENCES

- Borkowski, J.G., Benton, A.L. and Spreen, O., Word fluency and brain damage. Neuropsychologia, 5, 135-140, 1967.
- DeRenzi, E. and L.A. Vignolo, The Token Test: A sensitive test to detect receptive disturbances in aphasics. Brain, 85, 665-678, 1962.
- Goodglass, H. and Kaplan, E., The Assessment of Aphasia and Related Disorders. Philadelphia: Lea and Febiger, 1983.
- Howes, D., The word frequency concept in aphasia. In DeReuck and O'Conner (Eds.), Ciba Foundation Symposium on Disorders of Language. Baltimore: Little, Brown, 1964.
- Jones, L.V. and Wepman, J.M., A Spoken Word Count. Chicago: Language Research Associates, 1966.
- Kaplan, E., Goodglass, H. and Weintraub, S., Boston Naming Test. Philadelphia: Lea and Febiger, 1983.
- Porch, B.E., Porch Index of Communicative Ability. Palo Alto: Consulting Psychologists Press, Inc., 1967.
- Wertz, R.T., Word fluency measure. In F.L. Darley (Ed.), Evaluation of Appraisal Techniques in Speech and Language Pathology. Reading, Mass.: Addison-Wesley Publishing Company, 1979.
- Wertz, R.T. and Lemme, M.L., Input and output measures with aphasic adults. Research and Training Center 10, Final Report, Social Rehabilitation Services, Washington, D.C., 1974.
- Wertz, R.T., Collins, M., Weiss, D., Brookshire, R.H., Friden, T., Kurtzke, J.F. and Pierce, J., Veterans Administration Cooperative Study on Aphasia: A comparison of individual and group treatment. Journal of Speech and Hearing Research, 24, 4, 580-594, 1981.

DISCUSSION

Q: Did you collect word fluency data on other types of letters?

A: Yes. On three other letters.

Q: Well, it appeared that perhaps your Broca's may have had difficulty following the phonologic strategy for lexical access. That is, they seem to be down on the "S" in general and they did not use sound-link strategies in word retrieval. We looked at "S," "T," "P," and "F" in anterior and posterior patients, and found that the apraxic patients were significantly down on the phonological letter cue. More than that, they had more errors on fricative sounds than on stop sounds. In retrieving the words, they were able to retrieve fewer fricative words than stop words.

A: I don't think we have the data that you're looking for.

Q: Did you look at patient reliability on a test-retest basis?

A: No.

Q: Do you know of anybody who's done that on the Word Fluency Measure?

A: No.

C: We looked at "S," "T," "P," and "C," and found no group differences among the letters, which kind of surprised us. We thought we would. Another thing that kind of surprised me, if I remember the data correctly, is that your Broca's patients, these telegraphic, agrammatic patients, produced the largest number of adjectives compared with the other two groups.

A: That's right.

C: We found that the Word Fluency Measure correlated very highly with education, both in normals and aphasic patients. There is a correlation with age for young children, but not in the middle or older ages. It didn't correlate with the Coloured Progressive Matrices. The way people produce words seems to be a function of severity. A lot of folks give you no words, and you have to be a fairly mild to moderate aphasic patient to produce many words on this task at all. But the mild to moderate will tend to give you words across the minute of production, where the more severe, if they give you any words, will produce them early.

C: We correlated severity as measured by the PICA with some of these variables, and the ones that we looked at had a zero correlation almost across the board. So severity as measured by the PICA didn't seem to relate to what we looked at.