Word Choice in the Connected Speech of Aphasic and Non-brain-damaged Speakers

Robert H. Brookshire and Linda E. Nicholas

This study of aphasic speakers' word choice was part of a larger study of connected speech in aphasia. A major concern of that research is whether different eliciting stimuli lead to different patterns of word choice by speakers. This study is concerned with the effects of eliciting stimuli upon *vocabulary.* Schuell, Jenkins, and Jimenez-Pabon (1965) asserted that available vocabulary is always reduced in aphasia, an assertion that seems to be supported in studies by Fillenbaum, Jones, and Wepman (1961), Howes (1964), and Wachal and Spreen (1973). These studies suggest that aphasia restricts access to the less frequently occurring words in the lexicon.

It seems reasonable that different eliciting stimuli might lead to different patterns of word choice by aphasic speakers. For example, talking about one's own life and experiences might lead a speaker to use more common words than when describing pictures or telling stories. Data reported by Hayes (1988) support such an expectation. According to Hayes, normal speakers' word choices are affected by the situation in which the speech occurs; when they converse with other adults their word choice is different from when they converse with children, and their word choice in conversations differs from their word choice in interviews or courtroom testimony.

Accordingly, we designed a study to address the following questions:

—Do elicitation conditions that are typically used to elicit connected speech from aphasic adults result in patterns of word choice comparable to word choice in daily-life conversations?

—Do different elicitation conditions elicit different patterns of word choice?

—Do speech samples produced by aphasic speakers contain more common (high frequency) words and fewer uncommon (low fre-

quency) words than speech samples produced by non-brain-damaged speakers?

—Does severity or type of aphasia affect an aphasic speaker's word choice?

METHOD

Speech samples were obtained from 20 non-brain-damaged adults, 5 adults with nonfluent (essentially Broca's) aphasia, and 10 adults with fluent-mixed aphasia (fluent speech with literal paraphasias and word-retrieval difficulty). All were native speakers of English with adequate hearing and vision for the tasks. Non-brain-damaged subjects were non-hospitalized and non-institutionalized adults with a mean age of 64.2 years (SD = 7.3, range = 49–73) and a mean of 12.8 years of education (SD = 2.6, range = 6–17). Aphasic subjects were at least 3 months post onset of a single left-hemisphere cerebrovascular brain injury. The severity of aphasia was estimated by subjects overall percentile on a four-subtest short version (SPICA) (DiSimoni, Keith, & Darley, 1980) of the *Porch Index of Communicative Ability* ([PICA] Porch, 1971). Descriptive information for the aphasic subjects is given in Table 1.

The speech samples were obtained in response to the following stimuli:

—the picture from the *Boston Diagnostic Aphasia Examination* (Goodglass & Kaplan, 1983).

—the picture from the *Western Aphasia Battery* (Kertesz, 1982).

—two single pictures, each of which depicted a humorous situation in daily life

—two sequences of pictures, each of which depicted a humorous sequence of events in daily life

—two requests for descriptions of procedures:
"Tell me how you would go about writing and sending a letter."
"Tell me how you would go about doing dishes by hand."

—two requests for personal information:
"Tell me what you usually do on Sundays."
"Tell me where you live and describe it to me."

Subjects were tested individually in a quiet room. Each test session began with a short interval of practice and instruction to familiarize subjects with what they would be expected to do. Then the stimulus pictures

TABLE 1. CHARACTERISTICS OF APHASIC SPEAKERS

	Age	*Education*	*Months PO*	*SPICA %ile*
		Nonfluent (n = 5)		
Mean	58.0	14.8	141.0	70.6
SD	5.4	3.4	46.7	9.6
Range	51–65	12–20	71–199	59–85
		Fluent-Mixed (n = 10)		
Mean	66.5	12.6	36.4	64.6
SD	6.2	1.7	41.9	15.7
Range	54–76	11–16	4–109	40–82
		Total Group (n = 15)		
Mean	63.7	13.3	71.3	66.6
SD	7.1	2.5	66.0	13.9
Range	51–76	11–20	4–199	40–85

Note: Months PO = months post onset; SPICA %ile = overall percentile. Based on a short version of the Porch Index of Communicative Ability.

and requests were presented one at a time in random order. If the eliciting stimulus was a picture or set of pictures, the appropriate pictorial stimulus was placed on the table in front of the subject, and the subject was instructed to talk about it for at least one minute. If the eliciting stimulus was a request, a card containing the printed request was placed on the table in front of the subject and the request was read aloud by the examiner. The picture or card was left on the table until the subject finished talking about it. If a subject stopped talking before producing 15 seconds of speech, the examiner asked, "Can you tell me more?"

Subjects' responses to the stimuli were tape-recorded and orthographically transcribed by a typist. The reliability of the transcriptions was assessed by a second person who listened to the recording of each speech sample and compared it with its transcript. Instances in which the listener disagreed with the transcript were evaluated by the two investigators and a consensus decision was made regarding whether (and how) the transcript should be corrected.

Analyses of word choice were carried out with a computerized program that does lexical analysis of texts called QLEX (Hayes, 1989). QLEX, developed at Cornell University, Ithica, New York, determines how often each of the 10,000 most common word types in English occurs in a text. The word frequency norms come from the *American Heritage* corpus (Carroll, Davies, & Richman, 1971). Hayes (1988, 1989) has published word-frequency statistics for a variety of spoken and printed materials. Figure 1 shows a graph of word choice, based on information from Hayes.

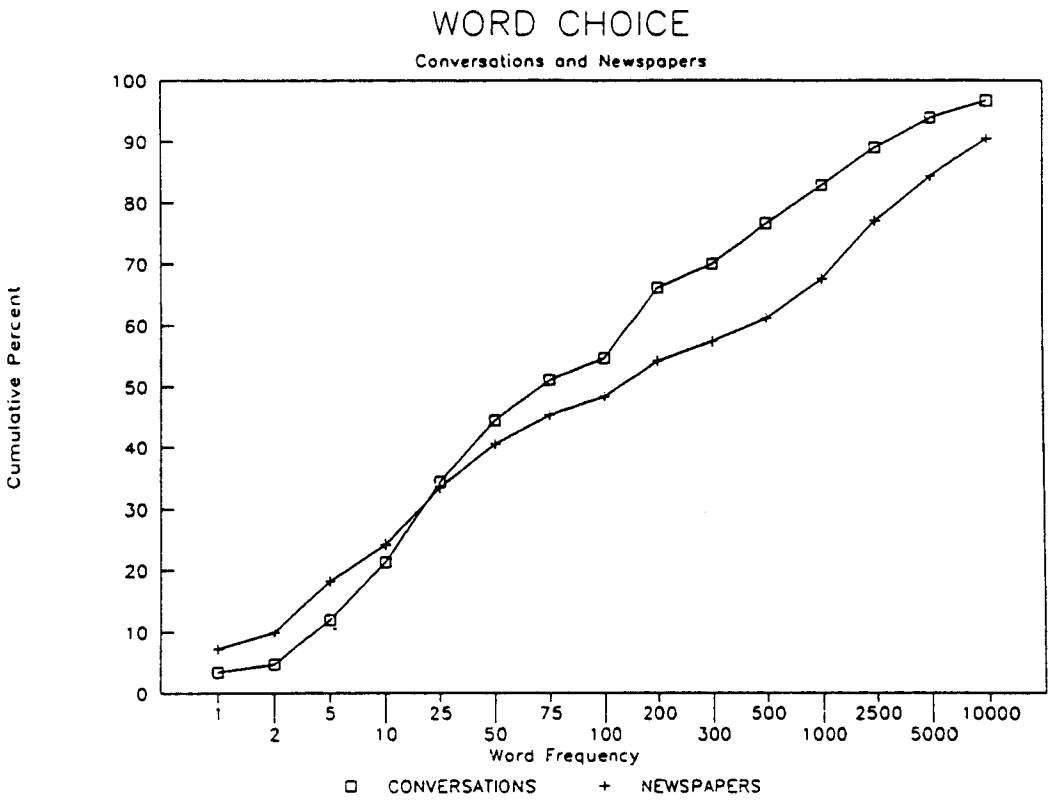

Figure 1. Patterns of word choice within adult-to-adult conversations and newspaper texts. (From data provided in Hayes, 1989.)

A line running from the lower left to the upper right depicts the pattern of word choice. The line represents the cumulative proportion of word types lying at or below word ranks from the most frequent (1, which is the word "the," at the left), to the 10,000th most frequent word type, at the right. The intercept of the right end of the line with the vertical axis is the proportion of the sample's word types that came from the 10,000 most common types in the *American Heritage* corpus. The farther the end of the line is from 100%, the more rare words—that is, words with a frequency of less than 1 in 10,000—there are in the sample. If a sample contains fewer rare words, its word-choice curve is displaced to the left and its endpoint is higher. One of Hayes's major findings was that common printed materials, such as newspapers and magazines, yield a linear function in which the proportions of words in the sampled materials almost exactly match the proportions in the *American Heritage* corpus, which was based on printed materials, but that spoken materials, such as adult-to-adult conversations, yield an S-shaped curve. The curve for adult-to-adult conversations is displaced to the left, relative to the curve for newspapers, and its endpoint is higher, suggesting, not surprisingly, that conversations contain

more common words and fewer rare words than newspaper texts. The 10,000 most frequent words account for about 97% of the words in adult-to-adult conversations, but only about 90% of the words in newspaper texts. According to Hayes, the initial bend in the S-curve for conversations is caused by speakers' systematic underuse of the first six most common words in English (the, of, and, a, to, in), and overuse of the next four (is, you, that, it) relative to printed texts.

Hayes found that in normal adult-to-adult conversations the 75 most common words account for approximately 50% of the words used. The 75 most common words in English are primarily (88%) function words, and the ranks above 75 are almost all (>99%) content words. The 1,000 most common words account for 83% of normal adult-to-adult conversations, and the first 5,000 account for 94%. By the time the 10,000th most common word is reached, 97% of the words in normal conversations are accounted for.

RESULTS

To determine whether word choice in speech samples obtained with the five kinds of eliciting stimuli resembles word choice in daily-life conversations, we compared what the non-brain-damaged speakers said in response to the five elicitation conditions with what Hayes's normal speakers said in adult-to-adult conversations. The comparison of what the non-brain-damaged speakers said in response to the pictures from the *Western Aphasia Battery* and the *Boston Diagnostic Aphasia Examination,* with Hayes's data for conversations and newspapers is illustrated in Figure 2a.

Non-brain-damaged speakers' word choice in talking about the aphasia test pictures more closely resembled Hayes's results for adult-to-adult conversations than for newspapers. This was also true for the other four kinds of eliciting stimuli (Figures 2b–2e).

However, the non-brain-damaged speakers tended to overuse the most common word types, compared with Hayes's adults. This tendency was greatest when they were describing single pictures (Figures 2a and 2b).

Table 2 compares the overall proportion of words accounted for at selected ranks for our non-brain-damaged speakers and Hayes's adults.

The proportions of the samples accounted for at each frequency rank were similar across elicitation conditions, and closely resembled those for Hayes's adults, except for procedural discourse, which elicited a slightly greater proportion of the 75 most frequent words than the other eliciting stimuli did.

These results also relate to the second question addressed in this study: whether different elicitation conditions elicit different patterns of word

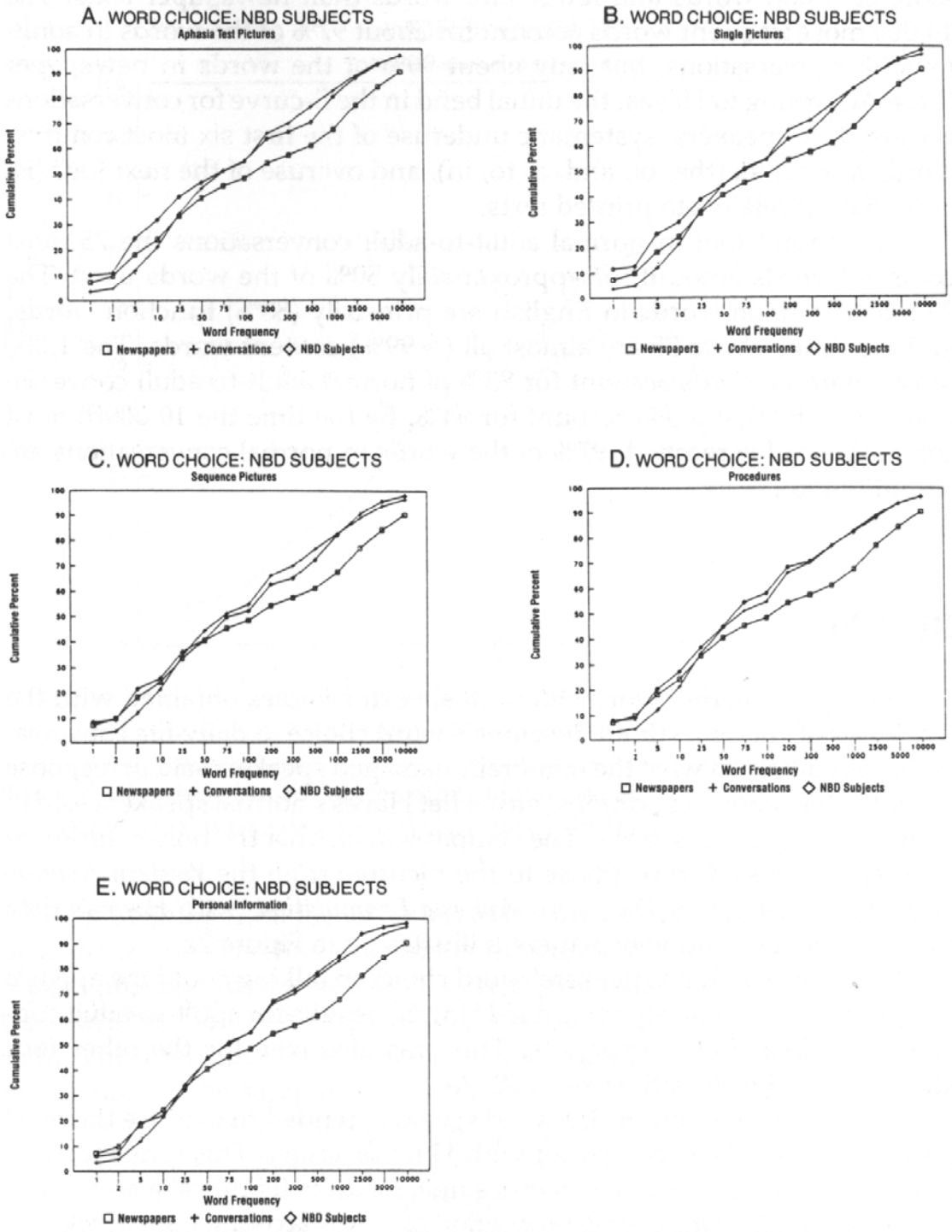

Figure 2. Word choice of non-brain-damaged speakers responding to five kinds of eliciting stimuli.

choice. The preceding graphs of word choice and Table 2 suggest that they do not, at least not from non-brain-damaged speakers.

Figure 3a shows the word frequency curve for aphasic speakers describing the two aphasia test pictures.

TABLE 2. OVERALL PROPORTIONS OF WORDS ACCOUNTED FOR AT EACH FREQUENCY RANK FOR HAYES'S ADULT-TO-ADULT CONVERSATIONS AND FOR NON-BRAIN-DAMAGED SPEAKERS RESPONDING TO FIVE ELICITING STIMULI

Freq	*CNV*	*ATP*	*SPIX*	*SQPX*	*PRO*	*PINF*
75	51	52	51	50	55	51
1,000	83	80	83	82	82	85
5,000	94	93	95	96	93	97
10,000	97	98	99	99	97	98

Note: CNV = Hayes's adult-to-adult conversations; ATP = aphasia test pictures; SPIX = single pictures; SQPX = picture sequences; PRO = procedures; PINF = personal information.

The curve for aphasic speakers resembles that for Hayes's adult-to-adult conversations, and it also resembles the curve for non-brain-damaged speakers, except for a more pronounced upward bow on the left. Thereafter, aphasic speakers slightly underuse word ranks 25 to 75, which brings the curve for our aphasic speakers and the curve for Hayes's adults together, after which they closely approximate each other. The same upward bow on the left is seen on the curves for the other eliciting stimuli (Figures 3b–3e). The bow is most pronounced for single pictures and least obvious when speakers provide personal information.

In order to determine the source of the bow, we did a word-by-word analysis of our aphasic speakers' use of the five most frequently occurring word types (*the, of, and, a,* and *to*). The aphasic speakers used the word *and* much more than Hayes's adult speakers did. The word *and* accounted for less than 3% of the words used by Hayes's adult speakers, but it accounted for 8% of the words produced by the fluent-mixed aphasic speakers, and for more than 14% of the words produced by the nonfluent aphasic speakers. This overuse of *and* crossed all elicitation conditions and appears to reflect aphasic subjects' propensity to use the word as an all-purpose continuant and filler, rather than as a conjunction. The non-brain-damaged speakers also showed the same tendency. "And" accounted for about 7% of their words, versus 3% for Hayes's adults.

Table 3 shows the cumulative proportions of word types at representative frequency ranks for the fluent-mixed aphasic speakers. The five stimuli elicited relatively consistent patterns of word choice from fluent-mixed aphasic speakers. Like non-brain-damaged speakers, fluent-mixed aphasic speakers relied on the 75 most frequent types more when they described procedures than when responding to the other stimuli. The aphasic speakers' tendency to overuse the most frequent word types, relative to Hayes's adults, is also evident in Table 3.

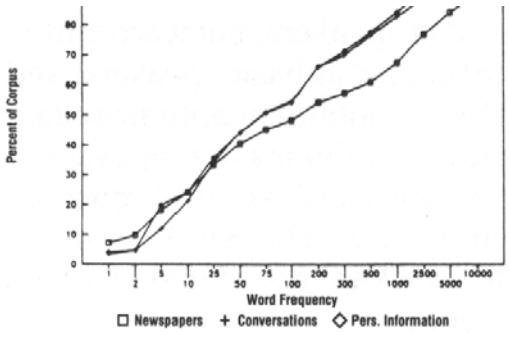

Figure 3. Word choice of aphasic speakers responding to five kinds of eliciting stimuli.

Table 4 presents the results for the nonfluent aphasic speakers. Nonfluent aphasic speakers exhibited somewhat more variability across eliciting stimuli than the other two groups, especially for word ranks 1 to 1,000. However, these data are for only five subjects, and the speech sam-

TABLE 3. OVERALL PROPORTIONS OF WORDS ACCOUNTED FOR AT EACH FREQUENCY RANK FOR HAYES'S ADULT-TO-ADULT CONVERSATIONS AND FOR MIXED-FLUENT APHASIC SPEAKERS RESPONDING TO FIVE ELICITING STIMULI

Freq	*CNV*	*ATP*	*SPIX*	*SQPX*	*PRO*	*PINF*
75	51	55	54	52	59	54
1,000	83	84	84	85	87	88
5,000	94	96	96	97	97	97
10,000	97	99	99	99	99	98

Note: CNV = Hayes's adult-to-adult conversations; ATP = aphasia test pictures; SPIX = single pictures; SQPX = picture sequences; PRO = procedures; PINF = personal information.

TABLE 4. OVERALL PROPORTIONS OF WORDS ACCOUNTED FOR AT EACH FREQUENCY RANK FOR HAYES'S ADULT-TO-ADULT CONVERSATIONS AND FOR NONFLUENT APHASIC SPEAKERS RESPONDING TO FIVE ELICITING STIMULI

Freq	*CNV*	*ATP*	*SPIX*	*SQPX*	*PRO*	*PINF*
75	51	43	48	48	51	45
1,000	83	83	83	78	71	77
5,000	94	94	94	93	90	93
10,000	97	97	97	98	95	96

Note: CNV = Hayes's adult-to-adult conversations; ATP = aphasia test pictures; SPIX = single pictures; SQPX = picture sequences; PRO = procedures; PINF = personal information.

ples from these subjects were shorter than those for fluent-mixed aphasic subjects so that perceived variability may be artifactually higher. The most striking finding was that the speech samples of the nonfluent aphasic subjects contained much smaller proportions of the 75 most common word types than Hayes's samples, as well as the samples from our non-brain-damaged and fluent-mixed aphasic speakers. This seems to reflect nonfluent aphasic speakers' well-known tendency to leave out function words, because the first 75 types are mainly function words.

To determine whether aphasia causes speakers to use more high-frequency words and fewer low-frequency words than normal speakers, we calculated the average proportion of words in four word-rank intervals across elicitation stimuli for the three groups in this study and compared the proportions with those from Hayes's adult-to-adult conversations. Table 5 shows the results of these calculations.

TABLE 5. PROPORTIONS OF SPEECH SAMPLES ACCOUNTED FOR AT SELECTED INTERVALS FOR HAYES'S ADULT-TO-ADULT CONVERSATIONS (CNV) AND FOR NON-BRAIN-DAMAGED (NBD), FLUENT-MIXED APHASIC (FM APH), AND NONFLUENT APHASIC (NF APH) SPEAKERS

Interval	*CNV*	*NBD*	*FM Aph*	*NF Aph*
1–75	51	52	55	47
76–1000	32	30	31	31
1001–5000	11	13	11	17
5001–10000	3	3	2	4

Table 5 shows that there is no dramatic shift toward more frequent words for aphasic speakers compared with non-brain-damaged speakers, except for a slight overuse of the 75 most frequent types by fluent-mixed aphasic speakers and slight underuse of those types by nonfluent aphasic speakers.

To determine if the severity of aphasia affected aphasic speakers' word choice, we calculated correlations between each aphasic speaker's overall SPICA percentile and his or her median word frequency for content words across elicitation conditions (the 75 most frequent words were excluded). For the 10 fluent-mixed aphasic subjects, the correlation coefficient was .25, suggesting a weak relationship between aphasia severity and median word frequency. For the five nonfluent aphasic patients the correlation coefficient was .51, suggesting a moderate relationship between aphasia severity and median word frequency.

CONCLUSIONS

The results of this study permit several conclusions.

—Stimuli typically used to elicit connected speech from aphasic persons appear to yield valid estimates of word choice in daily-life conversations. Non-brain-damaged adults' word choice when responding to these stimuli is equivalent to that in "normal" adult-to-adult conversations, except for slightly greater reliance on the five most common word types, especially the word "and."

—The different eliciting stimuli evaluated in this study appear to provide equivalent estimates of speaking vocabulary. This was true for both non-brain-damaged and aphasic speakers. The greatest variability across eliciting stimuli occurs within the 1,000 most common words for nonfluent aphasic subjects.

—Aphasia does not seem to reduce the availability of less common word types in any striking manner. Most of the differences between aphasic and non-brain-damaged speakers appear to be accounted for by nonfluent aphasic subjects' overuse of the word "and" as a continuant, together with their underuse of other function words.

—Aphasia severity, as measured by the four-subtest SPICA, is not strongly related to word choice in connected speech.

ACKNOWLEDGMENTS

This research was supported by the Research Service, Minneapolis Veterans Affairs Medical Center, and by VA General Medical Research.

REFERENCES

Carroll, J. B., Davies, P., & Richman, B. (1971). *The American Heritage word frequency book.* New York: Houghton-Mifflin.

DiSimoni, F. G., Keith, R. L., & Darley, F. L. (1980). Prediction of PICA overall score by short versions of the test. *Journal of Speech and Hearing Research, 23,* 511–516.

Fillenbaum, S., Jones, L. V., & Wepman, J. M. (1961). Some linguistic features of speech from aphasic patients. *Language and Speech, 4,* 91–102.

Goodglass, H., & Kaplan, E. (1983). *The Boston Diagnostic Aphasia Examination.* Boston: Lea & Febiger.

Hayes, D. P. (1988). Speaking and writing: Distinct patterns of word choice. *Journal of Memory and Language, 27,* 572–585.

Hayes, D. P. (1989). *Guide to the lexical analysis of texts.* Technical Report Series 89–96. Ithaca, NY: Cornell University, Department of Sociology.

Howes, D. (1964). Application of the word-frequency concept to aphasia. In A. V. S. de Reuck & M. O'Connor (Eds.), *CIBA Foundation Symposium on Disorders of Language.* Boston: Little-Brown.

Kertesz, A. (1982). *The Western Aphasia Battery.* New York: Grune & Stratton.

Porch, B. E. (1971). *The Porch Index of Communicative Ability.* Palo Alto, CA: Consulting Psychologists Press.

Schuell, H., Jenkins, J. J., & Jimenez-Pabon, E. (1965). *Aphasia in Adults.* New York: Harper & Row.

Wachal, R. S., & Spreen, O. (1973). Some measures of lexical diversity in aphasic and normal language performance. *Language and Speech, 16,* 169–181.