A System for Scoring Main Concepts in the Discourse of Non-brain-damaged and Aphasic Speakers

Linda E. Nicholas and Robert H. Brookshire

About 6 years ago, when we first became interested in evaluating clinically relevant changes in the informativeness of aphasic adults' connected speech with treatment, we found that the available discourse measures did not fit our needs. The measure that came closest was published in 1980 by Yorkston and Beukelman. They asked 78 non-brain-damaged adults to describe the Cookie Theft picture from the Boston Diagnostic Aphasia Examination (BDAE; Goodglass and Kaplan, 1983). From these descriptions they compiled a list of content units, which were bits of information mentioned by at least one of their non-brain-damaged subjects.

Because we wanted to evaluate the informativeness of connected speech in response to a number of different stimulus materials, we attempted to develop a reliable rule-based system for differentiating words that would be informative to a listener from those that would not be. We called this system correct information unit (CIU) analysis (Nicholas & Brookshire, 1988). A CIU is a word that is intelligible in context, accurate in relation to the eliciting stimulus, and relevant to and informative about the eliciting stimulus. (Words do not have to be used grammatically to be included in the CIU count.) After revising and expanding the rules for this measure for each of the studies in which we have used it (Brenneise-Sarshad, Nicholas, & Brookshire, 1991; Correia, Brookshire, & Nicholas, 1990; MacLennan, Nicholas, Morley, & Brookshire, 1992; Potechin, Nicholas, & Brookshire, 1987; Schumacher & Nicholas, 1991), we know that these rules can be used with high interjudge reliability to score connected speech from a wide range of aphasic speakers who are talking about a variety of different eliciting stimuli. We have found that two measures, percent of words that are CIUs and number of CIUs per minute, are sensitive to differences between non-brain-damaged and aphasic speakers and to changes in connected speech with recovery from aphasia.

As we worked with these two measures in various studies, we realized that there were still some important aspects of the informativeness of connected speech that we had not captured. One aspect relates to the number of main concepts (or how much of the gist) about a picture or topic the speaker conveys.

Something similar to what we are calling main concepts has been evaluated in the single-picture descriptions of aphasic adults (Gleason, Goodglass, Obler, Green, Hyde, & Weintraub, 1980; Ulatowska, Freedman-Stern, Doyle, & Macaluso-Haynes, 1983), in the Cookie Theft picture descriptions of adults with either aphasia or Alzheimer's disease (Heir, Hagenlocker, & Shindler, 1985; Nicholas, Obler, Albert, & Helm-Estabrooks, 1985), and in the descriptions of a picture sequence by right-brain-damaged adults (Joanette, Goulet, Ska, & Nespoulous, 1986). Each of these studies reported that, to varying degrees, the brain-damaged group produced fewer main concepts than the control group.

These studies did not report the procedures used to establish a list of main concepts for a stimulus or the procedures used to score the presence of main concepts in sufficient detail to allow others to use this type of analysis, nor did they report the test-retest or interjudge reliability of these judgments. Also, they made no attempt to evaluate the accuracy and completeness with which main concepts were produced by brain-damaged adults.

This study addressed the following questions:

—Can the presence, completeness, and accuracy of main concepts in discourse be scored with high interjudge reliability?

—Are the presence, completeness, and accuracy of main concepts stable across discourse samples elicited from the same subject on different occasions?

—Does the discourse produced by non-brain-damaged adults differ from that of aphasic adults in the presence, completeness, and/or accuracy of main concepts?

—Are measures of the presence, completeness, and/or accuracy of main concepts sensitive to changes in aphasic adults' connected speech with recovery?

METHOD

Subjects

Subjects were 20 non-brain-damaged adults and 15 aphasic adults. All were native speakers of English who demonstrated adequate hearing and

TABLE 1. DESCRIPTIVE INFORMATION FOR APHASIC SUBJECTS

	Age	*Education*	*Months PO*	*SPICA %ile*
		Nonfluent (n = 5)		
Mean	58.0	14.8	141.0	70.6
Std Dev	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
Std Dev	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
Std Dev	7.1	2.5	66.0	13.9
Range	51–76	11–20	4–199	40–85

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

vision for the tasks. Aphasic subjects were at least three months post onset of a single left-hemisphere cerebrovascular brain injury. Five exhibited nonfluent (essentially Broca's) aphasia and 10 exhibited fluent-mixed aphasia (fluent speech with literal paraphasias and word retrieval difficulty). The severity of aphasia was estimated by the overall percentile on a shortened version of the Porch Index of Communicative Disorders ([SPICA]; DiSimoni, Keith, & Darley, 1980). Descriptive information for aphasic subjects is summarized in Table 1. Non-brain-damaged subjects were non-hospitalized and non-institutionalized adults who were similar to aphasic adults in age (M = 64.2 years; SD = 7.3; range = 49–73) and education (M = 12.8; SD = 2.6; range = 6–17).

Stimulus Materials

The speech elicitation stimuli were selected from a larger study of discourse in aphasia of which this study is one part. They consisted of two single pictures from standard aphasia tests—the BDAE and the Western Aphasia Battery (WAB; Kertesz, 1982)—with two single pictures and two picture sequences that were drawn for the discourse production study, and two requests for procedural information: "Tell me how you go about doing dishes by hand" and "tell me how you go about writing and sending a letter." Figures 1a and 1b contain one of the single pictures and one

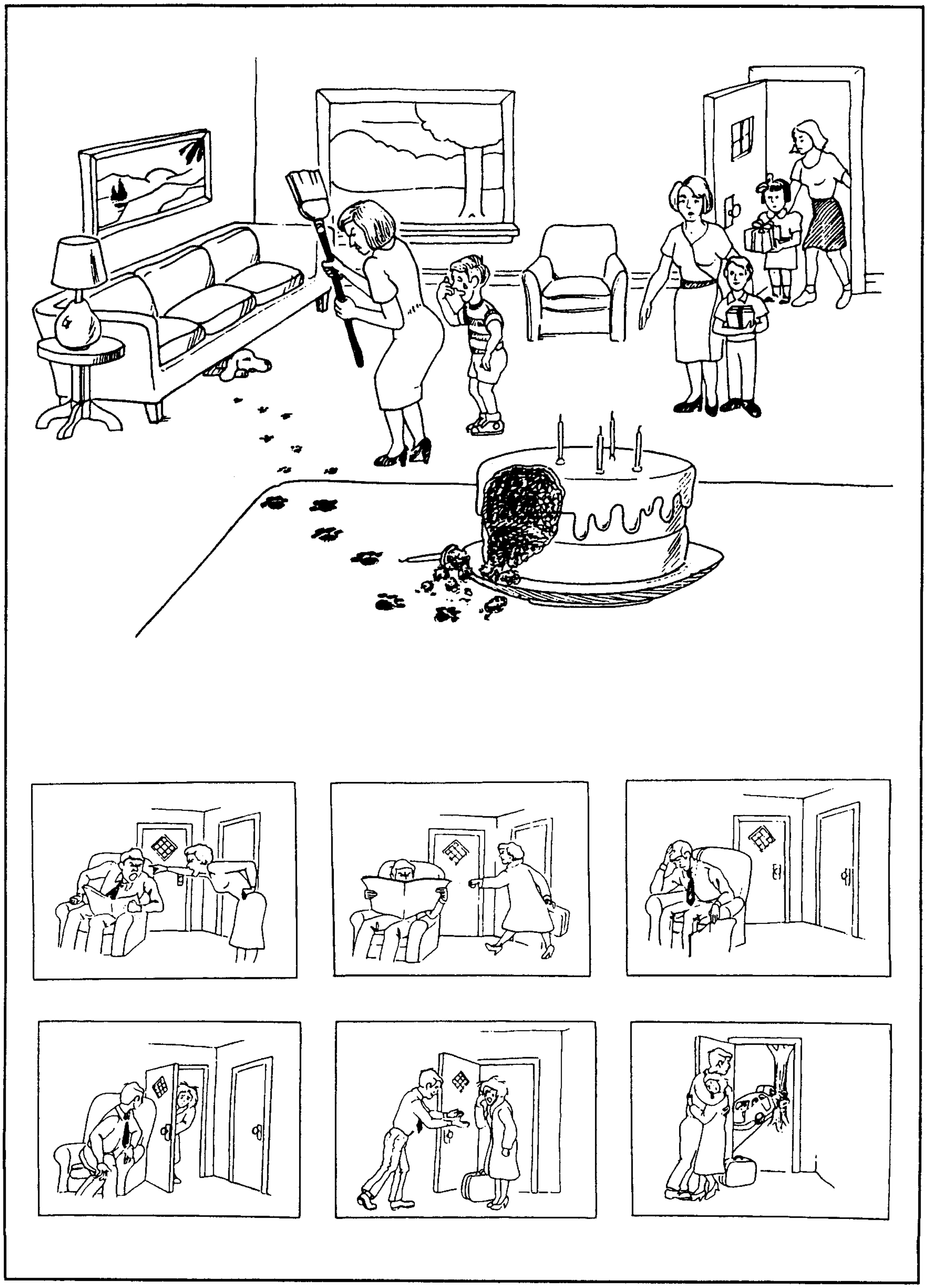

Figure 1a. (top) A single picture (The Birthday Party) and **Figure 1b. (bottom)** a picture sequence (The Argument) drawn for the discourse production study. (Copyright 1987 by R. Brookshire and L. Nicholas).

of the picture sequences that were drawn for the discourse production study.

Procedures

Following practice items and instruction, the stimulus pictures and procedural requests were presented individually to subjects in random order. The pictures and written procedural information requests were placed on a table in front of the subject and left there until the subject had completed the description. The requests for procedural information also were read aloud to the subject. Subjects were asked to try to talk about each of the elicitation stimuli for about a minute. If a subject stopped talking before producing 15 seconds of speech, they were prompted once with "Can you tell me more?" Each subject described the pictures and procedures in the same order three times. The first two sessions occurred on the same day and were separated by a 10-minute break. The third session occurred 7 to 10 days following the first session. Subjects were asked to make their descriptions as similar as possible across the three sessions. Subjects' descriptions were audiotaped and orthographically transcribed.

Validation of Main Concepts

Ten speech-language pathologists read the rules for writing main concept statements and completed several training tasks. Following feedback on their training-task performance, they were asked to write a list of main concepts for each of the stimulus items. Main concepts that were written in a similar form by 7 of the 10 judges were put on a list of main concept statements for each stimulus.

Next, the first author and a graduate student used these lists of main concepts and written rules and examples to identify and score main concepts in the Session 2 transcripts of the 20 non-brain-damaged subjects. Each main concept was bracketed on the transcript. Then the two scorers assigned one of the following five scores to each bracketed main concept:

—AC: accurate, complete

—AI: accurate, incomplete

—IN: inaccurate, complete

—II: inaccurate, incomplete

—AB: absent

TABLE 2. MAIN CONCEPTS FOR THE BIRTHDAY PARTY PICTURE

1. There is a birthday party (birthday cake).
2. The dog got into (messed up, took a bite of) the cake.
3. The woman (mother) is after the dog.
4. She has a broom.
5. The boy is crying.
6. People (guests) are arriving.

Main concepts that were present in some form in the transcripts of 14 of the 20 non-brain-damaged subjects (70%) were put on the final lists of main concepts statements. Interjudge reliability for determining whether main concept statements were present or absent was calculated on the transcripts of 12 randomly selected subjects. Reliability was above 90 percent for every subject, and overall reliability was 97%.

The final lists of main concepts for the eight stimuli contained 53 main concepts. These lists were used to score main concepts in 840 transcripts (35 subjects × 3 sessions × 8 stimuli). This was done by the same two scorers. The list of main concepts for the Birthday Party picture is in Table 2. The most common alternate words or information are in parentheses. Scorers also had an extensive list of examples with additional alternate wordings that convey the same basic idea for each concept.

RESULTS AND DISCUSSION

Interjudge Reliability

Point-to-point interjudge reliability was calculated for scoring main concepts in the transcripts of 12 non-brain-damaged and 12 aphasic subjects (one randomly selected session each). Reliability was calculated for two scorers. The first scorer was the graduate student who participated in the validation portion of the study. Because this scorer had been involved in that phase of the study and had helped with clarification of scoring rules and compiling scoring examples, it did not seem that her reliability scores would be a legitimate test of what interjudge reliability would be for someone who had not worked extensively with the scoring system. Therefore, another graduate student also scored the same 24 subjects' transcripts. She was given the written scoring rules and examples and several practice transcripts to score. After discussion of her scored practice transcripts, only minimal clarification was provided to this scorer.

TABLE 3. POINT-TO-POINT INTER-JUDGE RELIABILITY FOR SCORING MAIN CONCEPTS

	Mean	*Range*
Across NBD Subjects	93.9	86.8–100
Across APH Subjects	85.4	73.6–98.1
Across Stimulus Items	89.6	84.7–92.9

Note: NBD = non-brain-damaged. APH = aphasic.

The reliability data in Table 3 are for the second scorer; the means and ranges are nearly identical to those of the first scorer. Reliability was at acceptable levels for all eight stimulus items and for all non-brain-damaged subjects. Reliability was unacceptably low for one aphasic subject (73.6%) who was the most impaired of the fluent-mixed aphasic subjects (SPICA OA = 40 percentile). Reliability was also unacceptably low for the scoring category of *inaccurate, incomplete* (II). If some information in a main concept is inaccurate, it is often difficult to tell if the other information is complete. This scoring category was dropped and the scoring category of *inaccurate, complete* (IN) was changed to *inaccurate* with no specification of completeness. This was done before the second reliability check and all main concepts that had received a score of *inaccurate, incomplete* were rescored. This change resulted in a less symmetrical but more reliable scoring system.

Stability

Table 4 gives means, standard deviations, and ranges for percent of main concepts that received each of the four scores and one combination of scores (IN + AI) across the three sessions. Mean scores for both the non-brain-damaged and aphasic groups were fairly stable across the three sessions, with the smallest differences occurring between Sessions 2 and 3.

The actual differences in mean percentage for each of the scores across the three sessions are listed in Table 5. For non-brain-damaged subjects as a group, no average change exceeded 5.3%. Scores were most stable from Sessions 2 to 3, with no average change exceeding 1%. Scores from Sessions 1 to 2 and 1 to 3 increased in the percent of *accurate, complete* scores and decreased in percentages for the other three scores. For aphasic subjects as a group, no average change exceeded 2.4%. Their scores were also more stable from Sessions 2 to 3, with no average change exceeding 1.1%. The pattern of changes in scores from Sessions 1 to 2 and 1 to 3 were generally the same as those of the non-brain-damaged subjects.

TABLE 4. PERCENT MAIN CONCEPT SCORES ACROSS THREE SESSIONS

Score	Session 1			Session 2			Session 3		
	MEAN	SD	RANGE	MEAN	SD	RANGE	MEAN	SD	RANGE
				Non-brain-damaged (n = 20)					
AB	14.3	8.3	0–28	11.0	8.1	0–28	11.6	7.7	2–25
IN	2.4	2.8	0–11	1.6	1.5	0–6	1.1	1.3	0–4
AI	8.6	7.0	0–25	7.4	5.0	2–17	8.3	5.7	2–19
IN+AI	11.0	7.6	0–26	9.0	5.2	2–19	9.4	6.3	2–21
AC	74.7	14.4	51–96	80.0	11.8	59–96	79.0	13.3	57–94
				Aphasic (n = 15)					
AB	23.7	14.7	6–62	22.8	15.2	0–57	22.4	14.7	6–59
IN	10.4	8.1	2–32	9.2	8.1	0–30	9.8	6.4	0–28
AI	22.9	7.8	4–38	23.8	9.7	8–42	22.4	8.8	9–34
IN+AI	33.3	9.9	17–49	33.0	11.3	13–51	32.2	9.9	17–49
AC	43.0	19.5	6–68	44.3	23.5	8–79	45.4	20.7	8–74

Note: AB = Absent; IN = Inaccurate; AI = Accurate, incomplete; AC = Accurate, complete.

TABLE 5. MEAN PERCENT CHANGE IN MAIN CONCEPT SCORES ACROSS THREE SESSIONS

Group	Score	Sessions		
		1 TO 2	2 TO 3	1 TO 3
NBD	AB	-3.3	+0.6	-2.7
	IN	-0.8	-0.5	-1.3
	AI	-1.2	+0.9	-0.3
	IN+AI	-2.0	+0.4	-1.6
	AC	+5.3	-1.0	+4.3
APH	AB	-0.9	-0.4	-1.3
	IN	-1.2	+0.6	-0.6
	AI	+0.9	-1.4	-0.5
	IN+AI	-0.3	-0.8	-1.1
	AC	+1.3	+1.1	+2.4

Note: NBD = Non-brain-damaged (n = 20); APH = Aphasic (n = 15)
AB = Absent; IN = Inaccurate; AI = Accurate, incomplete; AC = Accurate, complete
+ = An increase in the number of main concepts receiving a score and - = A decrease.

TABLE 6. NUMBER (PERCENT) OF APHASIC SUBJECTS IN NON-BRAIN-DAMAGED SUBJECTS' RANGE FOR MAIN CONCEPT SCORING CATEGORIES

Scoring Category	*Session 2*	*Session 3*
AB	11 (73)	11 (73)
IN	8 (53)	4 (27)
AI	6 (40)	5 (33)
AC	6 (40)	6 (40)
AC+AI	8 (53)	9 (60)
AI+IN	3 (20)	3 (20)
AC+AI+IN	10 (67)	10 (67)

Note: AB = Absent; IN = Inaccurate; AI = Accurate, incomplete; AC = Accurate, complete.

Sensitivity

To determine if the presence, completeness, and/or accuracy of main concepts distinguished the discourse of aphasic speakers from that of non-brain-damaged speakers, the number and percent of main concepts that received each of the four scores or combinations of scores were calculated for the 20 non-brain-damaged subjects. Then the number and percent of aphasic subjects' score counts that fell into the non-brain-damaged subjects' range were tallied (Table 6).

Previous studies have found that brain-damaged subjects, as a group, failed to mention more main concepts than non-brain-damaged subjects. Our group data also show this pattern (Table 4). However, 73 percent of the aphasic subjects fell within the non-brain-damaged subjects' range for number of absent main concepts (Table 6). What best discriminated the performance of aphasic subjects from non-brain-damaged subjects was the number of main concepts that were judged to be *accurate, incomplete,* together with those that were judged to be *inaccurate* (AI + IN). In other words, it was not the number of main concepts that were completely left out that distinguished aphasic speakers from their non-brain-damaged counterparts, but rather the completeness and accuracy with which they produced the main concepts that they did mention.

To determine whether this measure of the presence, completeness, and accuracy of main concepts responds to changes in aphasic adults' connected speech production with recovery, we tracked the performance of several aphasic adults over time with the main concept measure and two other discourse measures (percent CIUs and CIUs per minute). We also tested these subjects with the *Boston Naming Test* (BNT; Kaplan, Goodglass, & Weintraub, 1983) and the repetition and naming subtests of the

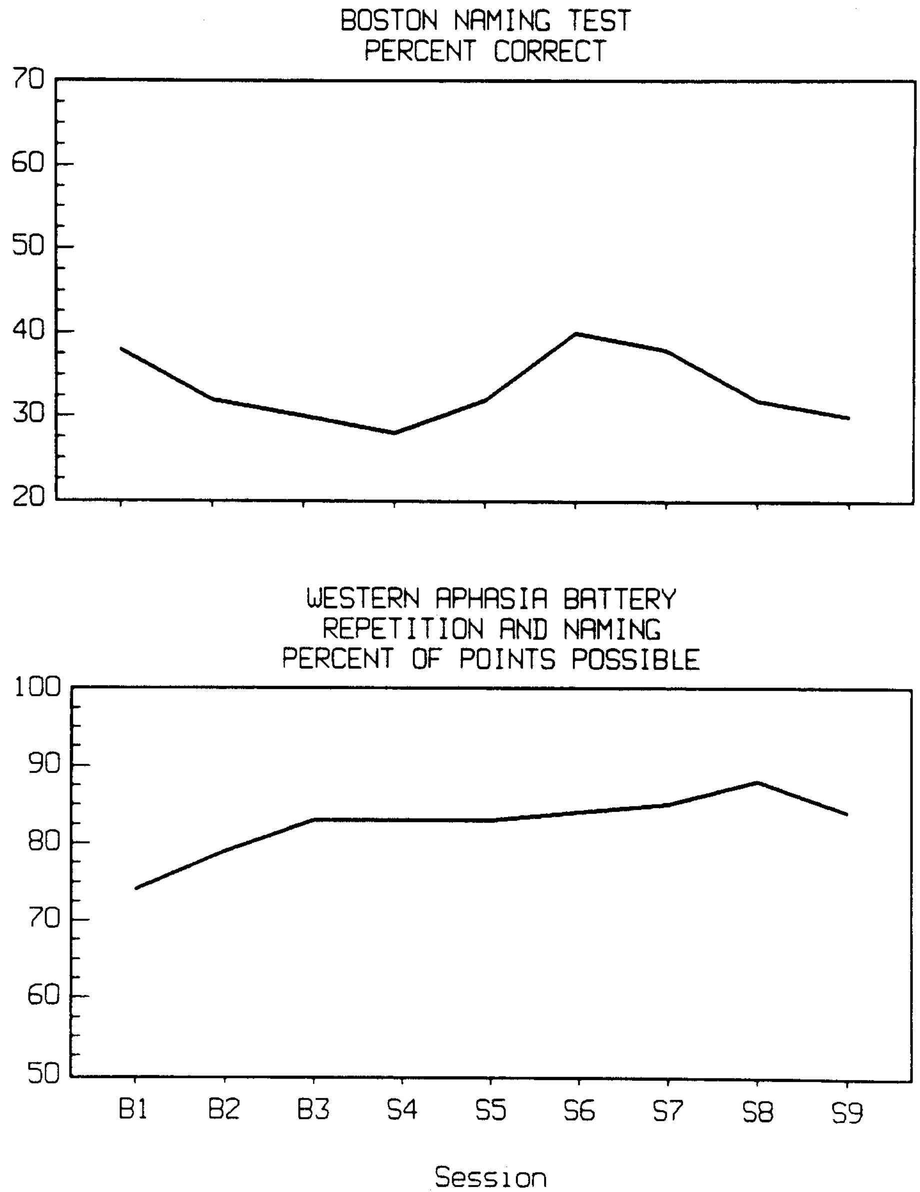

Figure 2. Performance across three baseline sessions (B) and six test sessions (S) on two standard measures of speech production by an aphasic adult.

WAB. The data in Figures 2 and 3 are for a man with fluent-mixed aphasia. Initially, his speech was quite empty with many literal and verbal paraphasias and occasional neologisms. Collection of speech samples and test data was initiated when he was one month post onset of his aphasia (at which time he participated in three baseline sessions on three consecutive days) and was continued every two weeks for the next three months. Figure 2 shows that his performance on the BNT and on the WAB repetition and naming subtests did not change appreciably over time from his

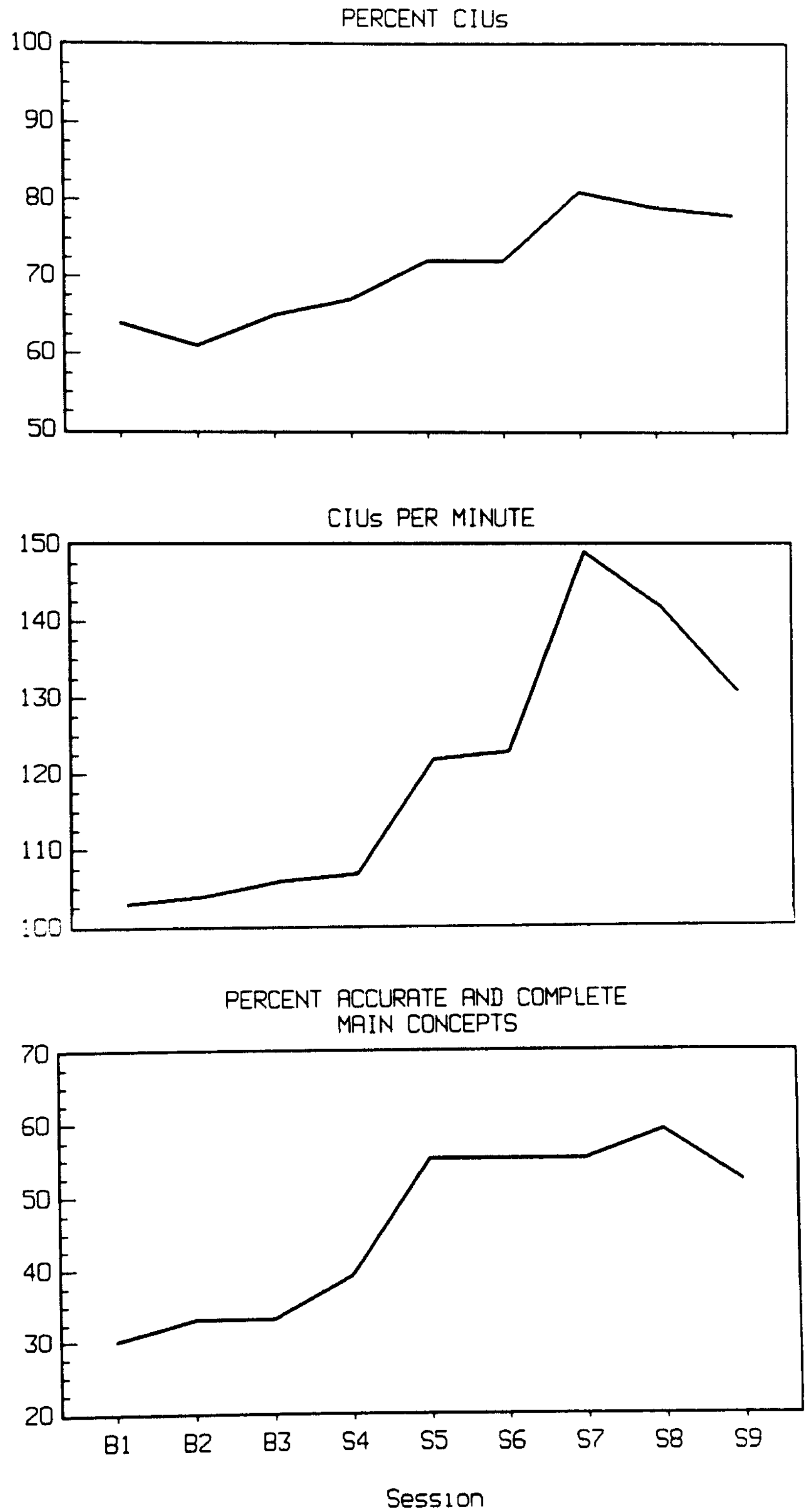

Figure 3. Performance across three baseline sessions (B) and six test sessions (S) on three measures of discourse production by an aphasic adult (CIUs = correct information units).

performance during the three baseline sessions. This sharply contrasts with his improved performance beyond the baseline phase on the three discourse measures shown in Figure 3.

In conclusion, this measure of the presence, completeness, and accuracy of main information in discourse appears to provide clinically relevant information about aphasic adults' production of connected speech. This measure can be scored with high interjudge reliability for aphasic adults with mild to moderately severe aphasia. Performance on this measure appears to remain relatively stable when no change should be occurring, with greatest stabilization taking place after the speaker's first experience with the eliciting stimuli. If information from this measure is combined with information about the rate and efficiency with which a speaker produces informative words, a picture begins to develop of the strengths and weaknesses that aphasic adults demonstrate when they produce connected speech. Such a combination of measures may permit a sensitive analysis of changes in connected speech with treatment.

ACKNOWLEDGMENTS

The authors wish to thank Shelley Brundage and Julia Edgar for their assistance with this research. This research was supported by the Department of Veterans Affairs Medical Research Service.

REFERENCES

Brenneise-Sarshad, R., Nicholas, L. E., & Brookshire, R. H. (1991). Effects of apparent listener knowledge and picture stimuli on aphasic and non-brain-damaged speakers' narrative discourse. *Journal of Speech and Hearing Research, 34,* 164–176.

Correia, L., Brookshire, R. H., & Nicholas, L. E. (1990). Aphasic and non-brain-damaged adults' descriptions of aphasia test pictures and gender-biased pictures. *Journal of Speech and Hearing Disorders, 55,* 713–720.

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

Gleason, J. B., Goodglass, H., Obler, L., Green, E., Hyde, M. R., & Weintraub, S. (1980). Narrative strategies of aphasic and normal-speaking subjects. *Journal of Speech and Hearing Research, 23,* 370–382.

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

Heir, D. B., Hagenlocker, K., & Shindler, A. G. (1985). Language disintegration in dementia: Effects of etiology and severity. *Brain and Language, 25,* 117–133.

Joanette, Y., Goulet, P., Ska, B., & Nespoulous, J. (1986). Information content of narrative discourse in right-brain-damaged right-handers. *Brain and Language, 29,* 81–105.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *The Boston Naming Test.* Philadelphia: Lea & Febiger.

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

MacLennan, D. L., Nicholas, L. E., Morley, G. K., & Brookshire, R. H. (1992). The effects of Bromocriptine on speech and language function in a man with Transcortical Motor Aphasia. In T. E. Prescott (Ed.), *Clinical Aphasiology,* (Vol. 20, pp. 145–156). Austin, TX: PRO-ED.

Nicholas, L. E., & Brookshire, R. H. (1988). *Correct Information Unit Analysis.* Minneapolis, MN: Aphasia Research Unit, Veterans Administration Medical Center.

Nicholas, M., Obler, L. K., Albert, M. L., & Helm-Estabrooks, N. (1985). Empty speech in Alzheimer's disease and fluent aphasia. *Journal of Speech and Hearing Research, 28,* 405–410.

Potechin, G. C., Nicholas, L. E., & Brookshire, R. H. (1987). Effects of picture stimuli on discourse production by aphasic and non-brain-damaged adults. In R. H. Brookshire (Ed.), *Clinical Aphasiology: Conference Proceedings.* Minneapolis, MN: BRK Publishers.

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

Schumacher, J. G., & Nicholas, L. E. (1991). Conducting research in a clinical setting against all odds: An unusual treatment for fluent aphasia. In T. E. Prescott (Ed.), *Clinical Aphasiology* (Vol. 19, pp. 267–277). Austin, TX: PRO-ED.

Ulatowska, H. K., Freedman-Stern, R., Doyle, A. W., & Macaluso-Haynes, S. (1983). Production of narrative discourse in aphasia. *Brain and Language, 19,* 317–334.

Yorkston, K. M., & Beukelman, D. R. (1980). An analysis of connected speech samples of aphasic and normal speakers. *Journal of Speech and Hearing Disorders, 45,* 27–36.