APHASIOLOGY, 2000, VOL. 14, NO. 5/6, 537–549

Linguistic validation of four parallel forms of a story retelling procedure

PATRICK J. DOYLE[1,2,3], MALCOLM R. McNEIL[1,2,3], GRACE PARK[1,3], AMY GODA[1], ELAINE RUBENSTEIN[4], KRISTIE SPENCER[1], BRIAN CARROLL[1,2], AMY LUSTIG[1,3], and LESLIE SZWARC[1,3]

[1] Aphasia Rehabilitation Research Laboratory & Clinic, VA Pittsburgh Healthcare System, USA
[2] Geriatric Research Education and Clinical Center (GRECC), VA Pittsburgh Healthcare System, USA
[3] Department of Communication Science and Disorders, University of Pittsburgh, USA
[4] Office of Measurement & Evaluation of Teaching, University of Pittsburgh, USA

Abstract

This study reports the development and validation of four parallel forms of a story retelling procedure. The equivalency of forms was based on the performance of 15 adults with aphasia on 12 operationally defined productive language variables including measures of (a) verbal productivity, (b) information content, (c) grammatical well-formedness, (d) phoneme production, and (e) verbal disruptions. The results revealed no significant differences among the four forms of the test for any of the dependent measures, and strong, positive and significant correlations among forms for 11 of the 12 dependent measures. These results suggest that a wide variety of productive language variables can be reliably measured using parallel forms of the story-retelling procedure described herein.

There is considerable evidence that several variables including characteristics of the elicitation stimuli, modality of stimulus presentation, and cognitive demands of the language task affect discourse performance in both non-brain-injured (Shadden *et al*. 1991, Wambaugh *et al*. 1991) and aphasic adults (Ulatowska *et al*. 1981; Bottenberg *et al*. 1987, Potenchin *et al*. 1987, Doyle *et al*. 1994; Doyle *et al*. 1995). Nevertheless, the precise way in which these variables affect subjects' performance, and how they may interact with patient characteristics such as the degree of working memory impairment or the severity of productive and/or receptive language processing deficits remains poorly understood.

Despite a relatively large body of literature addressing discourse performance in adults with aphasia, the relatively poor understanding of this area remains in part because there has been little systematic replication of the methods and measures used among the reported investigations. Rather, the literature is characterized by studies that have employed diverse stimuli and tasks to elicit discourse. Likewise, a variety of dependent variables have been used to describe connected speech samples obtained from persons

Address correspondence to: Patrick J. Doyle PhD, VA Medical Center, 7180 Highland Drive, Pittsburgh, PA 15206, USA. E-mail: patrick.doyle@med.va.gov

with aphasia. For example, Ulatowska *et al*. (1981) measured aspects of linguistic complexity and story grammar in self-generated narratives, picture descriptions, and story retellings of aphasic adults. Wambaugh *et al*. (1991) measured the proportionate use of speech acts of aphasic and normal adults under structured conversational and referential communication conditions. Saffran *et al*. (1989) used pictures to elicit a familiar story narrative (i.e. Cinderella) from which various morphological and sentential aspects of aphasic language were examined. Nicholas and Brookshire (1993) developed and standardized a discourse elicitation task that requires subjects to describe a set of six pictures, and to respond to two requests for personally relevant information, and two requests for procedural information. These investigators measured aphasic subjects' performance on several operationally defined measures of communicative informativeness and efficiency. Further, in a series of investigations examining the psychometric properties of the procedure, inter-rater reliability coefficients, standard errors of measurement for each dependent variable, test–retest stability based on discourse sample size in number of words, and cut off scores for 'normal' performance were reported (Nicholas and Brookhsire 1993, 1995, Brookshire and Nicholas 1994). The procedure has gained widespread clinical acceptance, and is being used with increasing frequency to assess changes in the connected speech of persons undergoing specific productive language interventions (Spencer *et al*. in press, McNeil *et al*. 1997, McNeil *et al*. 1998).

Although the procedures employed by Nicholas and Brookshire (1993, 1995) sample the range of narrative tasks that have been shown to influence discourse performance in persons with aphasia, they are substantively different from story-retelling tasks. Specifically, the former require the speaker to select lexical items and formulate story elements into a coherent narrative, whereas the latter require the speaker to retain story elements and their temporal order, retrieve these elements from memory, and reformulate them linguistically. Despite these differences and their potential influence on performance, very few studies have made direct comparisons across narrative and story-retelling conditions (Ulatowska *et al*. 1981, Shadden *et al*. 1991). Further, the results of these studies are difficult to interpret due to inherent differences in the nature and complexity of the stimuli employed across tasks.

More recently, Doyle *et al*. (1998) developed a discourse elicitation procedure using a standardized and well controlled set of stimuli (Brookshire and Nicholas 1993) that may be administered as a story-retelling procedure under picture-supported or oral-only stimulus elicitation conditions. This procedure permits direct comparisons of individual subjects' productive discourse performance under conditions of varying cognitive demands while controlling for the nature and complexity of the eliciting stimuli. In addition, because repeated measurement of discourse performance is frequently necessary in both clinical and research settings, and because repeated exposure to the same set of assessment stimuli may result in unwanted learning effects, the development of equivalent forms of the procedure was considered to be of great utility. The purpose of this investigation was to examine the extent to which the 12 individual stories comprising the Doyle *et al*. (1998) story-retelling procedure could be grouped to comprise equivalent forms of the instrument. Specifically, this investigation describes the development of four parallel forms of the story-retelling procedure based on the performance of 15 adults with acquired aphasia on 12 operationally defined productive language variables including measures of (a) verbal productivity, (b) information content, (c) grammatical well-formedness, (d) phoneme production, and (e) verbal disruptions (see Appendix).

Table 1. Descriptive subject information

Subjects	Age	MPO	RTT Percentile	ABCD Ratio	Raven's	PICA OA Percentile	PICA VRB Percentile
1	62	11	73	84.62	34	92	78
2	77	44	19	118.18	24	59	63
3	47	11	4	100	24	65	54
4	51	77	53	133.33	29	87	60
5	79	13	77	233.33	20	75	77
6	56	84	95	100	32	87	89
7	74	71	96	85.7	27	94	97
8	55	30	63	100	32	75	71
9	66	33	80	100	27	89	76
10	57	85	58	125	27	86	75
11	64	252	14	91	24	77	68
12	71	94	4	100	22	43	37
13	52	17	92	100	36	87	91
14	73	23	66	116.66	21	76	70
15	74	11	54	100	18	63	54
M	63.87	57.07	56.53	112.52	26.47	78.07	70.67
SD	10.45	62.12	32.10	36.15	5.33	14.90	15.71

MPO = Months post onset; RTT = Revised Token Test (McNeil and Prescott 1978), percentile compared to adults with left-hemisphere damage; ABCD ratio = Arizona Battery for Communication Disorders of Dementia (Bayles and Tomoeda 1993) ratio, determined by number of delayed recall items/number of immediate recall items × 10; Raven's = Raven's Coloured Progressive Matrices (Raven 1976), raw score out of a possible 36; PICA = Porch Index of Communicative Ability (Porch 1981), percentile compared to adults with left-hemisphere damage, OA = overall percentile and VRB = verbal percentile.

Method

Subjects

A total of 15 adults with aphasia due to a single left-hemisphere stroke participated in the investigation. All subjects were native speakers of English and passed a pure tone audiometric screening at 35dB HL at 500, 1000, 2000, and 4000 Hz unilaterally. The diagnosis of aphasia was based on clinical criteria operationalized from the formal definition of McNeil (1988), and was determined by clinical examination and formal testing conducted by the investigators. Descriptive information is displayed in table 1.

Stimuli

A total of 12 stories originally taken from the Discourse Comprehension Test (Brookshire and Nicholas 1993) served as stimuli for the story-retelling task. These stories are controlled for number of words, number of sentences, mean sentence length, number of subordinate clauses, number of T-units, ratio of clauses to T-units, listening difficulty and number of unfamiliar words. The stories were read and digitally recorded at a rate of 170 syllables/min by a male speaker seated in a double-walled sound booth. Each story was also illustrated by an artist as a six-plate black and white drawing. These drawings were then digitized. A PC-based computer program was developed to present the oral and pictured versions of the story in synchrony, and to record subjects' retelling of the story. Doyle *et al*. (1998) provides a sample story and its illustration.

Experimental procedures

Subjects were seated individually in a quiet room in front of a computer workstation with a 13-inch monitor and external speakers. Eye to monitor distance was held constant and the audio portion of each story was presented at 40dB HL above each subject's pure tone average at 1, 2, 3, and 4 KHz. Stories were presented in random order across subjects.

Following a set of brief instructions regarding the nature of the task, the pre-recorded oral versions of the stories were played while subjects viewed the monitor. As they listened, individual pictures comprising the six-plate sequence appeared on the screen in temporal correspondence with the oral version of the story. Immediately following the presentation of each story, all individual pictures appeared on the monitor as a six-plate picture sequence and subjects were instructed to use the pictures to retell the story in their own words. Subjects' retellings of each story were recorded online by the program for subsequent orthographic transcription and scoring. All 12 stories were presented within a single session of approximately 50 minutes duration.

Scoring and reliability measures

Recordings were orthographically transcribed into a microcomputer by trained research staff using transcription conventions described by Campbell and Dollaghan (1987). Transcripts were then scored using the operational definitions of each dependent variable. Measurement reliability data were obtained from two samples for each subject for a sampling rate of 17%. Point-to-point inter-rater reliability between two trained examiners was calculated by dividing the total number of agreements by the total number of agreements plus disagreements and multiplying by 100 for the following measures: utterance segmentation (100%), transcription of words and mazes (98%), correct information units (95%), story proposition accuracy and completeness (91%), syntactic errors (88%), sound production errors (95%), silent pauses greater than 2 seconds (87%).

Parallel forms development

To develop parallel forms, two goals were identified in arranging the 12 stories into four subsets of 3 stories. The first goal was to maximize consistency across the four subsets, or, in other words, to form the subsets so that correlations between all pairs of subsets should be high on the 12 variables. The second goal was to minimize differences among the four subsets with respect to measures of central tendency on the 12 variables.

The first step involved computing out a hierarchical cluster analysis among the stories with respect to the 12 variables. It would not have been possible to submit subject-level data to the cluster analysis procedure because of the unmanageable number of data points (12 variables $\times$ 15 subjects = 180 data points for each story). Instead, means and standard deviations of each of the 12 variables for each story were submitted to cluster analysis. Because of the large variability among units of measurement for the 12 variables, the means and standard deviations were standardized by conversion to z-scores.

Results of the cluster analysis revealed that stories 1, 10, 11, and 12 formed homogeneous groupings. In other words, among the 12 stories, these four stories were most similar to each other with respect to the 12 variables. Therefore, as a starting point in forming the subsets, it was decided to assign one of these four stories to each of four

subsets. The rationale was that this strategy would help to foster consistency among the subsets.

To assign the remaining eight stories to subsets, a balancing strategy was carried out based on examination of the stories standardized means on the 12 variables. For example, it was decided to assign stories 5 and 6 to the same subset because the z-score of mean for story 5 on percent accurate and complete story propositions was large and positive, whereas the z-score of the mean for story 6 was large and negative on the same linguistic variable. Similarly, it was decided to assign story 1 and story 4 to the same subset because the z-score of the mean for story 1 on number of mazes/minute was fairly large and negative, whereas the z-score of the mean for story 4 on number of mazes/minute was fairly large and positive. This process was continued for the other variables across the four forms.

As a result, homogeneous subsets of stories were grouped to yield four forms that were maximally similar. Specifically, stories 3, 8, and 10 comprised Form A; stories 5, 6, and 12 comprised Form B; stories 2, 7, and 11 comprised Form C; and stories 1, 4, and 9 comprised Form D. These forms were subsequently examined using non-parametric tests of differences, and rank order correlations to determine the extent to which the four forms yielded equivalent measures of each dependent variable. Alpha was set at $p < .01$ for all comparisons.

Results

Table 2 displays means and standard deviations for each dependent variable across the four forms of the sampling procedure. These data were submitted to Friedmans repeated measures ANOVAs on ranks to examine the magnitude of difference among test forms for each dependent variable. Table 3 shows the obtained Chi-square and probability values which revealed no significant difference among the four forms of the sampling procedure for any of the 12 dependent measures.

When non-significant findings are obtained and sample size is fairly small, it is not unreasonable to attribute the lack of statistical significance to insufficient power. While statistical significance is influenced by sample size, measures of effect size are a function only of the magnitude of the difference or strength of the relationship. Therefore measures of effect size are commonly calculated when results do not reach significance. To evaluate the magnitude of the difference among the parallel forms on the dependent variables, eta squared, a measure of effect size, was computed for each variable. The value of eta squared indicates the proportion of variance in the dependent variable that is explained by treatment, or in other words, the form of the test. Values of eta squared ranged from .009 to .0196 across the dependent variables, with a mean value of .09. Therefore, the maximum proportion of variation in performance explained by form of the test was less than 20% while the average proportion was less than 10%.

Table 4 displays Spearman rank order correlation coefficients among the four forms of the sampling procedure for each dependent variable. These data revealed strong, positive and significant correlations among test forms for 11 of the 12 dependent measures with rho values ranging from .61 to .99 ($p < .01$). Correlation among test forms for the measure *mean length of silent pause*, yielded moderate, and non-significant correlations for two comparisons [Forms A & B (rho = .54, $p = .04$) and A & D (rho = .56, $p = .03$)].

Standard errors of measurement were calculated for each dependent measure as a further estimate of the reliability with which the four forms of the sampling procedure measured performance on each dependent variable. SEM values were calculated using

Table 2. Means and (standard deviations) for verbal productivity, information content, grammatical structure, sound errors, and verbal disruptions across the four story forms

	A	B	C	D
Verbal Productivity				
Total words	408.20 (123.05)	423.93 (132.13)	435.00 (152.80)	382.53 (102.51)
Words/min	75.22 (33.13)	77.87 (32.85)	75.90 (34.86)	69.32 (30.90)
Mean length of utterance	10.64 (2.83)	11.14 (2.81)	11.05 (2.65)	10.40 (2.82)
Type Token Ratio	0.48 (0.08)	0.47 (0.08)	0.49 (0.09)	0.50 (0.08)
Information content				
Number of correct information units/min	41.61 (26.58)	44.88 (28.79)	43.39 (31.49)	42.65 (31.08)
% Correct information unit	54.55 (20.33)	56.80 (19.03)	56.69 (21.38)	56.18 (20.53)
% Accurate and complete story propositions	37.15 (24.71)	38.76 (22.51)	37.77 (29.50)	37.04 (29.55)
Grammatical Structure & Sound Errors				
% Grammatical well-formed utterances	76.20 (16.05)	79.24 (14.96)	77.22 (18.66)	73.07 (20.20)
Percent sound production errors	1.48 (1.70)	1.68 (1.76)	1.57 (1.88)	1.68 (2.28)
Verbal Disruptions				
Mazes/min	8.77 (3.87)	8.48 (3.93)	8.73 (3.20)	8.48 (3.38)
Silent pauses/min	0.67 (1.06)	0.63 (0.87)	0.67 (0.77)	0.68 (0.84)
Mean length of silent pause (in seconds)*	4.01 (1.30)	4.13 (0.49)	4.02 (0.80)	3.95 (1.09)

Story forms = A (3, 8, 10); B (5, 6, 12); C (2, 7, 11); D (1, 4, 9)
* Form A, n = 11; Form B, n = 9; Form C, n = 10; Form D, n = 11.

the following formula: SEM = SD $\sqrt{1 - \text{rho}}$, where the SD = $\sqrt{(\text{Var}_1 + \text{Var}_2)/2}$. These data are displayed in table 5 and reveal variability among the four forms of the test for several dependent measures but generally acceptable SEMs given the magnitude of the scores reported for each measure.

Discussion

The purpose of this study was to describe the linguistic development and validation of parallel forms of a story-retelling procedure by examining the performance of 15 adults with aphasia on 12 operationally defined measures of connected discourse. The findings revealed that subjects' performance did not differ significantly across the four forms of the story-retelling procedure for any of the 12 language production measures. However,

Table 3. Friedmans repeated measures ANOVA on ranks

	Chi-square	D.F.	Significance ($p < .01$)
Total words	8.84	3	0.03
Words/min	7.80	3	0.05
Mean length of utterance	7.38	3	0.06
Type token ration	6.90	3	0.08
Number of correct information units/min	4.60	3	0.20
% correct information units	2.28	3	0.52
% accurate and complete story propositions	1.80	3	0.62
% grammatical well-formed utterances	5.80	3	0.12
% sound production errors	1.82	3	0.61
Mazes/min	0.44	3	0.93
Silent pauses/min	0.50	3	0.92
Mean length of silent pauses	1.96	3	0.58

Table 4. Spearman rank correlation coefficients

	Form A	Form B	Form C
Verbal Productivity Measures			
Total Words			
Form B	.88		
Form C	.94	.93	
Form D	.88	.91	.90
Words/min			
Form B	.96		
Form C	.95	.99	
Form D	.94	.99	.98
Mean length of utterance			
Form B	.81		
Form C	.61	.76	
Form D	.79	.85	.64
Type Token Ratio			
Form B	.80		
Form C	.89	.86	
Form D	.93	.75	.89
Information Content Measures			
Number of correct information units/min			
Form B	.94		
Form C	.81	.89	
Form D	.81	.77	.84
% Correct information units			
Form B	.91		
Form C	.84	.80	
Form D	.79	.84	.91

Continued

Table 4. Continued

	Form A	Form B	Form C
% Accurate and complete story propositions			
Form B	.81		
Form C	.61	.76	
Form D	.79	.85	.64
% Grammatical well-formed utterances			
Form B	.86		
Form C	.91	.93	
Form D	.89	.84	.93
% Sound production errors			
Form B	.87		
Form C	.88	.87	
Form D	.93	.89	.95
Verbal Disruption Measures			
Mazes/min			
Form B	.93		
Form C	.95	.94	
Form D	.88	.91	.93
Silent pauses/min			
Form B	.77		
Form C	.78	.74	
Form D	.66	.74	.86
Mean length of silent pause			
Form B	.54*		
Form C	.77	.83	
Form D	.56*	.73	.90

Story forms = A (3, 8, 10); B (5, 6, 12); C (2, 7, 11); D (1, 4, 9). All correlation coefficients are positive and significant ($p<.01$) with the exception of those noted with * (Forms A & B, $p = .04$ and forms A & D, $p = .03$).

Table 5. Standard error of measurement for twelve linguistic variables across four story forms

	Form A	Form B	Form C
Verbal Productivity Measures			
Total Words			
Form B	44.23		
Form C	33.98	37.79	
Form D	39.23	35.48	41.14
Words/min			
Form B	6.60		
Form C	7.60	3.39	
Form D	7.85	3.19	4.66

Continued

Table 5. Continued

	Form A	Form B	Form C
Mean length of utterance			
Form B	1.23		
Form C	1.71	1.34	
Form D	1.29	0.84	1.64
Type Token Ratio			
Form B	0.04		
Form C	0.03	0.03	
Form D	0.02	0.03	0.03
Information Content Measures			
Number of correct information units/min			
Form B	6.79		
Form C	12.70	13.49	
Form D	12.61	14.37	12.51
% Correct information units			
Form B	5.91		
Form C	8.34	9.05	
Form D	9.36	7.92	6.29
% Accurate and complete story propositions			
Form B	10.30		
Form C	16.99	12.85	
Form D	12.48	10.17	17.71
% Grammatical well-formed utterances			
Form B	5.80		
Form C	5.22	6.10	
Form D	6.05	5.90	4.35
% Sound production errors			
Form B	0.62		
Form C	0.62	0.66	
Form D	0.53	0.68	0.47
Verbal Disruption Measures			
Mazes/min			
Form B	1.03		
Form C	0.79	0.88	
Form D	1.26	1.10	0.87
Silent pauses/min			
Form B	0.47		
Form C	0.43	0.42	
Form D	0.55	0.44	0.30
Mean length of silent pause			
Form B	1.23		
Form C	0.84	0.80	
Form D	1.13	0.98	0.58

Story forms = A (3, 8, 10); B (5, 6, 12); C (2, 7, 11); D (1, 4, 9).

a lack of significant difference does not necessarily indicate that the four forms were equivalent. The proportion of variance that was accounted for by the test form, however, does suggest that the non-significant differences were not attributable to the lack of statistical power or an effect size that was incapable of detecting a difference, but rather argues in favour of a lack of actual differences among the four forms. More importantly, correlation coefficients among forms were high and positive for 11 of the dependent variables including (a) total words, (b) words/min, (c) MLU, (d) TTR, (e) correct information units/min, (f) percent correct information units, (g) percent accurate and complete story propositions, (h) percent of grammatically well-formed utterances, (i) percent of sound production errors, (j) number of mazes/min, and (k) number of silent pauses/minute.

This procedure has several advantages over other discourse elicitation procedures described in the literature. First the stories comprising the instrument are controlled for a number of linguistic variables that may affect performance including the number of words, number of sentences, mean sentence length, number of subordinate clauses, number of T-units, ratio of clauses to T-units, listening difficulty and number of unfamiliar words. Second, pre-recorded pictured and oral stimuli ensures consistency of measurement conditions across individual stories comprising each form of the instrument, and across equivalent forms when used for repeated measurements. Third, the software operating the assessment tool allows selection of oral-only versus picture-supported stimulus presentation conditions, and thus permits direct comparison of performance using equivalent stimuli under varying task demand conditions. Fourth, unlike previously reported procedures in which the focus of measurement has been on single categories of linguistic behaviour such as information content (Nicholas and Brookshire 1993, 1995), morphology (Saffran *et al*. 1989), or story grammar (Ulatowska *et al*. 1981), the parallel forms of the current procedure have been validated on a variety of linguistic behaviours.

Although the results of this investigation should be viewed as preliminary and require replication with larger samples of persons with aphasia, they suggest that a wide variety of productive language variables can be reliably measured using the parallel forms of the story-retelling procedure described herein. Within the estimated standard error of measurement for each linguistic production variable, the four parallel forms of the story-retelling test can serve as a clinical or research tool for the quantification of language production in aphasia.

Acknowledgements

This study was supported by the Department of Veterans Affairs Rehabilitation Research & Development Project No. C894-2RA, and the VA Pittsburgh Healthcare System Geriatric Research, Education and Clinical Center.

References

Bayles, K.A., and Tomoeda, M.S. (1993) *Arizona Battery for Communication Disorders of Dementia*. (Tucson, Az: Canyonlands Publishing.)

Bottenberg, D., Lemme, M.L., and Hedenberg, N. (1987). Effect of story content on narrative discourse of aphasic adults. *Clinical Aphasiology*, *17*, 202–209.

Brookshire, R.H., and Nicholas, L.E. (1993). *Discourse Comprehension Test* (Tucson: Communication Skill Builders).

Brookshire, R.H., and Nicholas, L.E. (1994). Speech sample size and test–retest stability of connected speech measures for adults with aphasia. *Journal of Speech and Hearing Research*, *37*, 399–407.

Campbell, T.F., and Dollaghan, C.A. (1987). *Conventions for transcribing the spontaneous language samples of brain-injured children* (Children's Hospital of Pittsburgh Language Analysis Laboratory, Department of Communication Disorders).
Doyle, P.J., Goda, A.J., and Spencer, K.A. (1995). The communicative informativeness and efficiency of connected discourse by adults with aphasia under structured and conversational sampling conditions. *American Journal of Speech Language Pathology*, *4*, 130–134.
Doyle, P.J., McNeil, M.R., Spencer, K.A., Goda, A., Cottrell, K., and Lustig, A. (1998). The effects of concurrent picture presentation on retelling of orally presented stories by adults with aphasia. *Aphasiology*, *12*, 561–574.
Doyle, P.J., Thompson, C.K., Oleyar, K.S., Wambaugh, J.L., and Jackson, A.V. (1994). The effects of setting variables on conversational discourse in normal and aphasic adults. *Clinical Aphasiology*, *22*, 135–144.
McNeil, M.R. (1988). Aphasia in the adult. In: L.V. McReynolds, J. Northern, & D.E. Yoda (Eds.) *Handbook of speech-language pathology and audiology* (Philadelphia: W.B. Saunders).
McNeil, M.R., Doyle, P.J., Small, S., Spencer, K., Jackson-Goda, A., & Flores, D. (1997). A double blind, placebo controlled study of pharmacological and behavioral treatment of lexical semantic deficits in aphasia. *Aphasiology*, *11*, 385–400.
McNeil, M.R., Doyle, P.J., Spencer, K., Goda, A.J., and Flores, D. (1998). Effects of training multiple form classes on acquisition, generalization and maintenance of word retrieval in a single subject. *Aphasiology*, *12*, 575–585.
McNeil, M.R. and Prescott, T. (1978). *The Revised Token Test*. (Austin, TX: Pro-Ed).
Nicholas, L.E., and Brookshire, R.H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research*, *36*, 338–350.
Nicholas, L.E., and Brookshire, R.H. (1995). Presence, completeness, and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. *Journal of Speech and Hearing Research*, *38*, 145–156.
Porch, B.E. (1981). *Porch Index of Communicative Ability*. (Palo Alto, CA: Consulting Psychologists Press).
Potenchin, G.C., Nicholas, L., and Brookshire, R.H. 1987, Effects of picture stimuli on discourse production by aphasic speakers. *Clinical Aphasiology*, *17*, 216–220.
Raven, J.C. (1965). *Coloured Progressive Matrices*. (New York: The Psychological Corporation).
Saffran, E.M., Berndt, R.S., and Schwartz, M.F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language*, *37*, 440–479.
Shadden, B., Burnette, R., Eikenberry, B., and DiBrezzo, R. (1991). All discourse tasks are not created equal. *Clinical Aphasiology*, *20*, 327–341.
Shertzer, M. 1986, *The elements of grammar* (New York: Macmillan Publishing Co).
Spencer, K.A., Doyle, P.J., McNeil, M.R., Wambaugh, J.L., Park, G., and Carroll, B. (this issue). Examining the facilitative effects of rhyme in a patient with output lexicon damage. *Aphasiology*, *14*, 5/6.
Ulatowska, H.K., North, A.J., and Macaluso-Haynes, S. (1981). Production of narrative and procedural discourse in aphasia. *Brain and Language*, *13*, 345–371.
Wambaugh, J.L., Thompson, C.K., Doyle, P.J., and Camarata, S. 1991, Conversational discourse of aphasic and normal adults: An analysis of communicative functions. *Clinical Aphasiology*, *20*, 343–353.

Appendix

Measures of verbal productivity

1. *Total words:* Total number of words in each sample that were intelligible in context to someone who knows the topic/story being discussed (Nicholas and Brookshire, 1993).
2. *Words per minute:* Total number of words in the sample divided by the total duration of the sample in minutes.
3. *Mean length of utterance:* Average number of words per utterance; calculated by dividing the total number of words in the sample by the total number of utterances.

4. *Type token ratio:* The number of different words in the sample divided by the number of total words in the sample.

Measures of information content

5. *Number of correct information units/minute:* Total number of correct information units divided by the time (in minutes) taken to provide each story.
6. *Percent correct information units:* Total number of correct information units divided by the total number of words.
7. *Percent accurate and complete story propositions:* Total number of accurate and complete story propositions divided by the total number of story propositions.

CIUs are defined as words that are intelligible in context, accurate in relation to the story, and relevant to the content of the story. CIUs do not have to be used in a grammatically correct manner to be included in the CIU count. Each CIU consisted of a single word (Nicholas and Brookshire 1993).

Story propositions are the main ideas conveyed in each story. Accurate and complete story propositions were defined as propositions which contain essential components that must be included in the subject's retell of the story. Essential components of each story proposition were underlined in the listings of the story propositions. The wording of essential information did not have to be the same as that of the listed propositions, but the general meaning must have been the same. Essential information did not have to be given in standard grammatical form or standard word order, as long as deviations did not lead to miscomprehension of the essential meaning of the concept (adapted from Nicholas and Brookshire 1995).

Measure of grammatical structure[1]

8. *Grammatical well-formedness:* An overall measure of the accuracy and completeness of syntactic use which was calculated by dividing the number of accurate and complete clauses (including independent clauses, dependent clauses, and prepositional phrases) by the total number of clauses and phrases.

Analysis of syntactical structures was based on the delineation of story retells into the grammatical components of *independent clauses*, *dependent clauses*, and *prepositional phrases*. *Independent clauses* are defined as a group of words which has a subject and a predicate, and can stand alone as a sentence. *Dependent clauses* are a group of words which have a subject and a predicate, but cannot stand alone in a sentence. A dependent clause is incomplete and is used to modify and independent clause. Dependent clauses can be adjective clauses, adverb clauses, or noun clauses. Infinitive and participial phrases, which are similar in nature to dependent clauses, were coded as dependent clauses. (Infinitive phrases are the combination of an infinitive (combining 'to' with a verb) and an object. Participial phrases begin with either a present or past participle and are used as adjectives.) A *prepositional phrase* is a connecting word that shows the relation of a noun or a pronoun to other words in the sentence. Prepositional phrases are composed of a preposition, plus an object of the preposition and its modifiers. Prepositional phrases do not contain a subject or a predicate.

[1] Adapted in part from Shertzer 1986.

Measure of sound production

9. *Percent sound production errors:* Total number of syllables containing sound production errors divided by the total number of syllables produced in each story.

The following sound production errors were included: *additions* (the insertion of a vocalic or consonantal sound into a target syllable that does not belong in that syllable, *substitutions* (the replacement of a target sound by another sound from a different phonological sound class), *omissions* (the deletion of a target sound or syllable segment), and/or *distortions* (an approximation of the target sound which is in the same phonological class as the target sound).

Measures of verbal disruptions

10. *Mazes per minute:* The number of mazed utterances divided by the time taken to provide each story (in minutes).
11. *Pauses per minute:* The number of pauses greater than 2 seconds divided by the time taken to provide each story (in minutes).
12. *Mean length of pause time* (in seconds): Total duration of silent pauses greater than 2 seconds divided by the total number of pauses produced in each story.

Mazes were defined as repetitions (e.g. [I was] I was going), revisions (e.g. [He asked] He told the lady), word fragments (e.g. [bl] blue), and fillers (e.g. uh, um). Pauses, occurring either within or between utterances, were coded if they were greater than 2 seconds.