

Increasing the sensitivity of the Story Retell Procedure for the discrimination of normal elderly subjects from persons with aphasia

Malcolm R. McNeil and Patrick J. Doyle

*University of Pittsburgh, Passavants Hospital, and
VA Pittsburgh Health Care System, USA*

Grace H. Park, Tepanta R.D. Fossett, and Martin B. Brodsky

VA Pittsburgh Health Care System and University of Pittsburgh, USA

Background: Clinicians have long recognised the need for assessing language production at multiple levels of complexity and at impairment, participation, and activity levels. Methods for the elicitation of connected spoken language have taken many forms, typically selected with a balance between validly sampling linguistic performance, and reliability and economy of the sampling and scoring procedures. A Story Retell Procedure (SRP) has been proposed as a preferred method for achieving valid, reliable, and economic assessment of connected language (Doyle et al., 2000), and an information unit (IU) metric has been developed for validly and economically capturing important linguistic aspects of the retelling (McNeil, Doyle, Fossett, Park, & Goda, 2001).

Aims: In keeping with the goal of making assessment procedures as efficient and economic as possible, a study was undertaken to investigate the refinement of the IU metric for increasing the sensitivity of the SRP as an instrument for the detection of connected paragraph-level language production deficits in persons with aphasia. This metric involved the calculation of the percentage of IUs (%IU) produced relative to the time taken to produce them (%IU/Min).

Methods & Procedures: A total of 15 persons with aphasia, and 31 normal control individuals without a communication disorder served as participants for this study. Subjects heard, and immediately retold each of 12 stories originally taken from the Discourse Comprehension Test (Brookshire & Nicholas, 1997). The retellings were scored using the procedures outlined by McNeil et al. (2001) with the addition of the %IU calculated over the time of the retelling. Comparisons between subject groups and groups stratified by age, among SRP forms, between scoring methods (%IU vs %IU/Min.), and group misclassification by scoring method were made.

Outcome & Results: Application of the %IU/Min with the SRP yielded equivalence among alternate forms as evidenced by non-significant differences and high correlation coefficients among the SRP forms for persons with aphasia. The %IU/Min also decreased the percentage of misclassified aphasic and normal individuals compared to the %IU measure. Older normal subjects were misclassified as aphasic with greater frequency compared to the younger normal subjects.

Conclusions: The %IU/Min is a more sensitive metric than the %IU in differentiating individuals with aphasia from older normal controls.

Address correspondence to: Malcolm R. McNeil, PhD, Professor and Chair, Department of Communication Science and Disorder, University of Pittsburgh, 4033 Forbes Tower, Pittsburgh, PA 15260, USA. Email: McNeil+@pitt.edu

This material is based on work supported by the Office of Research and Development, Rehabilitation R&D Grant #B2265R, Department of Veterans Affairs, awarded by Drs McNeil and Doyle.

Due to the importance of evaluating connected spoken language in persons with aphasia (Nicholas & Brookshire, 1995; Ulatowska, Macaluso-Haynes, & North, 1980; Wambaugh, Thompson, Doyle, & Camarata, 1991), several procedures have been developed to elicit connected language production, including conversational observation (Oelschlaeger & Thorne, 1999), scripted interviews (Goodglass & Kaplan, 1983), on-line video narration (McNeil, Small, Masterson, & Fossett, 1995), off-line video scene description (Chafe, 1980), fable generation/story telling (Berndt, Wayland, Rochon, Saffran, & Schwartz, 2000; Ulatowska, Chapman, Highley, & Prince, 1998), picture description procedures (Nicholas & Brookshire, 1993, 1995; Yorkston & Beukelman, 1980), and procedural description (Nicholas & Brookshire, 1993, 1995). In addition to these methods, Doyle, McNeil, Spencer, Goda, Cottrell, and Lustig (1998) reported a story retelling procedure (SRP) that had the advantage over other elicitation procedures that language formulation and production occurred within a constrained context and therefore provided a standardised and replicable sampling procedure.

Following the report of the SRP's internal validity (Doyle et al., 1998) and high alternate-forms reliability (Doyle et al., 2000), a scoring metric was developed to quantify the informativeness of connected spoken language in individuals with aphasia as well as those with unimpaired linguistic ability (McNeil et al., 2001). The Percent Information Unit (%IU) was defined as "an identified word, phrase, or acceptable alternative from the story stimulus that is intelligible and informative and that conveys accurate and relevant information about the story" (McNeil et al., 2001, p. 994). Although high reliability and validity of the %IU metric were reported, the sensitivity of this metric to discriminate between individuals with aphasia and neurologically unimpaired persons appeared questionable (McNeil et al., 2001). Specifically, normal subjects were discriminated from persons with aphasia only 54% of the time on average across SRP forms. In addition, aphasic persons were differentiated from normal subjects with 77% average accuracy using 2 standard deviations \pm the standard error of measurement (SEM).

Following this study, and in the context of evaluating the memory requirements of the SRP task using the %IU metric, Brodsky et al. (2000), discovered that their normal young subjects performed significantly better than older normal subjects. This finding raised the possibility that the discriminative power of the SRP and the %IU metric might be different when persons with aphasia are compared to young normal individuals and when they are compared to older normal adults.

Finally, as the speed of cognitive processing has been reported to decrease with age (see review by Balota, Dolan, & Duchek, 2000), it was hypothesised that the sensitivity of the %IU metric and the SRP would be enhanced with the addition of an efficiency measure. That is, the percentage of information units produced per minute (%IU/Min) was hypothesised to provide a more sensitive measure as evidenced by a greater level of differentiation between older normal and aphasic persons than the %IU metric. Therefore, the purposes of this study were the following: (1) To compare older normal, younger normal, and aphasic adults' production of %IU/Min on the SRP; (2) to establish the equivalency of the four SRP forms for measuring %IU/Min in aphasic and normal adults; and (3) to compare the accuracy with which the %IU and the %IU/Min metrics discriminate between normal older, normal younger, and aphasic adults.

METHOD AND PROCEDURE

Participants

A total of 15 persons with aphasia and 31 normal individuals served as subjects. The subjects with aphasia were all English-speaking volunteers and ranged in age from 47 to 74 years ($M = 62.7$ years, $SD = 9.1$ years). These subjects were selected based on their performance within the range for left-hemisphere-damaged individuals, and by the demonstration of deficits across all modalities on the Porch Index of Communicative Ability (PICA; Porch 1981). The average aphasia severity, as measured by the PICA overall percentile, was 78 (range = 43–94, $SD = 14.3$). The normal subjects were between the ages of 23 and 80 years ($M = 43.7$ years; $SD = 17.2$ years) and were without self-reported history of neurological disease. All subjects passed a vision screening involving the reduced Snellen chart with a viewing distance that was equal to that of the computer screen in the experiment. Additionally, all subjects passed a hearing screening involving pure tones at 35 db HL at 0.5, 1, 2, and 4 kHz. Finally, all subjects obtained no greater than two points difference between the Immediate and Delayed Story Recall Tasks from the *Arizona Battery of Communication Disorders of Dementia* (Bayles & Tomoeda, 1993). Biographical and descriptive data for the aphasic subjects are presented in Table 1.

TABLE 1
Subject description

Subject	Age	PICA VRB percentile	PICA AUD percentile	PICA OA percentile	RTT (shortened) percentile
1	72	78	73	92	73
2	67	63	72	59	19
3	47	54	64	65	4
4	51	60	99	87	3
5	69	86	99	85	77
6	56	89	99	87	95
7	74	97	99	94	96
8	55	71	72	75	63
9	67	76	72	80	94
10	57	75	99	86	58
11	65	78	69	86	54
12	71	37	54	43	5
13	52	91	99	87	80
14	74	70	99	78	66
15	74	54	54	63	21
<i>M</i>	62.73	71.93	81.53	77.80	53.87
<i>SD</i>	9.14	16.19	17.85	14.30	34.50

PICA = *Porch Index of Communicative Ability* (Porch, 1981), percentile compared to adults with left hemisphere damage, VRB = verbal, AUD = auditory, OA = overall; RTT = *Revised Token Test* (McNeil & Prescott, 1978; Arvedson, McNeil, & West, 1985), percentile scores for adults with left-hemisphere damage.

Materials and procedures

Connected language samples were obtained from subjects using all four SRP forms under experimental conditions specified in Doyle et al. (2000). Retellings were orthographically transcribed and coded for %IUs using the operational definition and criteria specified in McNeil et al. (2001).

As the number of possible IUs was known from the stimulus stories, the number of IUs produced was divided by the number possible, to derive the %IUs. Subsequently, the %IUs for each story was divided by the total duration, or the total number of minutes calculated to the hundredths, for each subject's retelling of that story. As a result, a measurement of %IUs Per Minute (%IUs/Min) was obtained for each story and each subject. The %IUs/Min data were averaged across the three pre-selected stories for each SRP form.

RESULTS

Subject group comparisons

In order to evaluate the differences in %IU/Min production among the subject groups, separate ANOVAs were computed for each SRP form. These between-group effects were further examined using Student-Newman-Keuls pairwise multiple comparisons. The aphasic subjects produced significantly fewer ($p < .05$) %IUs/Min than all normal group comparisons for each of the SRP forms. The young and old normal groups did not differ significantly ($p > .05$) from each other on any SRP form.

Equivalency of SRP forms

In order to assess the equivalency of the SRP forms as a function of the %IUs/Min metric, repeated measures ANOVAs were conducted for the data obtained from the normal and aphasic subject groups. Results revealed a non-significant [$F(3) = 0.03$, $p = .99$] difference in the %IUs/Min produced among the four SRP forms for the aphasic group. However, significant SRP form effects were obtained for each of the three normal subject groups, the combined normal group [$F(3) = 30.74$, $p < .001$], old normal group [$F(3) = 8.33$, $p < .001$], and young normal group [$F(3) = 27.38$, $p < .001$]. For the combined normal subject group, post hoc pairwise multiple comparisons using the Student-Newman-Keuls test revealed significantly greater ($p < .05$) %IUs/Min in form B than for form A, as well as significantly greater ($p < .05$) %IUs/Min in SRP form D than in all other SRP forms. Similarly, for both the old and young normal subject groups, Student-Newman-Keuls tests revealed significantly greater ($p < .05$) %IUs/Min in SRP form D than in all other SRP forms.

To assess the degree of association between the %IUs/Min produced for each SRP form, a correlational analysis was also conducted. Across SRP forms, within-group Pearson product moment correlation coefficients were obtained for the aphasic and normal subject groups. Obtained correlation coefficients ranged from .96 to .97 for the aphasic group and from .75 to .93 for the three normal groups. Results are displayed in Table 2.

Subject group misclassification

Table 3 displays the mean, standard deviations, SEM, cutoff scores, percentage, and number of subjects misclassified for the four subject groups, for each of the four SRP forms and for the average across forms for the %IU measure. Table 4 displays the same form of the data for the %IU/Min measure.

TABLE 2
 Pearson product moment correlation coefficients for %IU/Min for the
 aphasic and normal subject groups

	<i>B</i>	<i>C</i>	<i>D</i>
<i>Aphasic subjects</i>			
A	.96	.96	.97
B		.96	.96
C			.97
<i>Normal subjects</i>			
A	.85	.91	.93
B		.77	.87
C			.88
<i>Old normal subjects</i>			
A	.81	.93	.96
B		.75	.85
C			.94
<i>Young normal subjects</i>			
A	.89	.89	.90
B		.81	.91
C			.83

All correlation coefficients are positive and significant ($p < .05$).

In order to identify the percentage of subjects misclassified based on %IU/Min between the aphasic and the three normal groups, cutoff scores were calculated for each subject group. For each of the young normal, old normal, and combined normal groups, a lower limit of performance was calculated using the group mean minus 2 SD. For the group of subjects with aphasia, an upper limit of performance was computed using the group mean plus 2 SD. The SEM for each SRP form was subtracted from each normal subject's %IU/Min score. For the aphasic group, the SEM for each SRP form was added to each subject's %IU/Min score. These SEM adjusted scores for each aphasic subject were then compared to the cutoff scores of each of the three normal groups. Similarly, the adjusted scores for the combined normal, young normal, and old normal subjects were compared to the aphasic cutoff score. Inspection of the data across forms and normal groups indicated a range of 13–27% misclassification of aphasic subjects, and 13–26%, 6–13%, and 20–47% misclassification of combined normal, young normal, and old normal subjects respectively. Number and percentage of misclassifications are presented in Table 4.

Comparison of %IU and %IU/Min misclassifications

Overall, the data presented in Table 3, compared to the data presented in Table 4, show a similar number of misclassifications for subjects with aphasia and substantively more, approximately twice as many, misclassifications of normal subjects using the %IU metric. Using either performance measure, the percentage of misclassified old normal subjects was greater than misclassified young subjects. Across SRP forms, 14% of young and 40% of old normal subjects were misclassified as aphasic using the %IU measure and only 8% and 32% young and old normal subjects respectively were misclassified using the %IU/Min metric.

TABLE 3
%IU

	A	B	C	D	Overall
<i>Aphasic (n = 15)</i>					
Mean,	22.04,	22.57,	22.61,	21.62,	
(SD),	(11.12),	(10.05),	(11.67),	(13.81),	
SEM,	3.21,	2.55,	2.66,	3.57,	
Cutoff score, (M + 2SD)	45.50,	42.74,	45.94,	49.24,	
% (Num) Misclassified as					
Combined normal	20, (3)	27, (4)	27, (4)	20, (3)	23, (14)
Young normal	20, (3)	13, (2)	7, (1)	7, (1)	12, (7)
Old normal	27, (4)	27, (4)	27, (4)	27, (4)	27, (16)
<i>Combined normal (n = 31)</i>					
Mean,	48.75,	50.59,	48.44,	54.30	
(SD),	(7.49),	(8.83),	(7.23),	(7.95)	
SEM,	3.78,	4.76,	3.61,	3.69	
Cutoff score, (M - 2SD)	33.77,	32.93,	33.98,	38.40	
% (Num) Misclassified as aphasic	55, (17)	36, (11)	55, (17)	39, (12)	46, (57)
<i>Young normal (n = 16)</i>					
Mean, (SD),	50.67,	54.26,	50.95,	57.11,	
SEM,	(7.16),	(7.98),	(5.01),	(7.41),	
Cutoff score, (M - 2SD)	3.58,	4.21,	2.81,	4.01,	
% (Num) Misclassified as aphasic	36.36,	38.29,	40.93,	42.30,	
	25, (4)	6, (1)	13, (2)	13, (2)	14, (9)
<i>Old normal (n = 15)</i>					
Mean, (SD),	46.70,	46.67,	45.73,	51.30,	
SEM,	(7.53),	(8.17),	(8.45),	(7.61)	
Cutoff score, (M - 2SD)	4.09,	5.02,	4.08,	3.46,	
% (Num) Misclassified as aphasic	31.64,	30.33,	28.33,	36.08,	
	47, (7)	40, (6)	40, (6)	33, (5)	40, (24)

Means, standard deviations, SEM, and cutoff scores for %IU, and percentages and number of subjects misclassified for the four subject groups and SRP forms and for the average across all forms.

DISCUSSION

This study reports the development of a metric designed to increase the sensitivity of the Story Retell Procedure for quantifying connected language samples based on the concept of communication efficiency. As previously stated, the purposes of this study were threefold. First, the production of %IU/Min was compared across the three normal and the aphasic subject groups. As expected, the aphasic subject group produced significantly fewer %IUs/Min than the young and old normal groups. However, given the findings that older normal subjects perform more poorly on various cognitive tasks (Carlesimo, Sabbadini, Fadda, & Caltagirone, 1997; Korsnes & Gilinsky, 1993; Korsnes & Magnussen, 1996), the non-significant difference between the old and young normal groups was unexpected.

The second purpose of this study was to compare the equivalency of the four SRP forms in the aphasic subject groups, as well as the three normal subject groups. Adding the dimension of time to the %IU metric previously described has yielded a measure that maintains the equivalence of the alternate forms of the SRP for persons with aphasia as evidenced by a lack of significant differences across forms and by high correlation

TABLE 4
%IU/Min

	A	B	C	D	Overall
<i>Aphasic (n = 15)</i>					
Mean, (SD),	13.1, (10.4),	13.1, (9.5),	13.3, (11.8),	13.0, (12.3),	
SEM,	4.9,	4.8,	5.1,	5.6,	
Cutoff score, (M - 2SD)	33.9,	32.2,	36.9,	37.6,	
% (Num.) Misclassified as					
Combined Normal	27, (4)	27, (4)	27, (4)	20, (3)	25, (15)
Young Normal	27, (4)	20, (3)	13, (2)	13, (2)	18, (11)
Old Normal	27, (4)	27, (4)	27, (4)	20, (3)	25, (15)
<i>Combined normal (n = 31)</i>					
Mean, (SD),	43.4, (9.8),	46.3, (11.3),	44.3, (9.9),	52.3 (11.4)	
SEM,	3.2,	4.7,	3.8,	3.7	
Cutoff score, (M - 2SD)	20.52,	19.03,	20.56,	25.77	
% (Num) Misclassified as aphasic	19, (6)	19, (6)	26, (8)	13, (4)	19, (24)
<i>Young normal (n = 16)</i>					
Mean, (SD),	44.7, (8.7),	48.1, (9.9),	45.1, (7.4),	54.8, (10.6),	
SEM,	2.9,	3.6,	2.9,	3.7,	
Cutoff score, (M - 2SD)	24.42,	24.76,	27.34,	29.89,	
% (Num) Misclassified as aphasic	13, (2)	6, (1)	6, (1)	6, (1)	8, (5)
<i>Old normal (n = 15)</i>					
Mean, (SD),	42.1, (11.0),	44.4 (12.8)	43.4, (12.3)	49.6, (12.0)	
SEM,	3.5,	5.7,	4.4,	3.5,	
Cutoff score, (M - 2SD)	16.58,	13.14,	14.41,	22.13,	
% (Num) Misclassified as aphasic	27, (4)	33, (5)	47, (7)	20, (3)	32, (19)

Means, standard deviations, SEM, and cutoff scores for %IU/Min, and percentages and number of subjects misclassified for the four subject groups and SRP forms and for the average across all forms.

coefficients among the forms. This suggests that the four SRP forms can be used equivalently or alternately to quantify the language production of persons with aphasia. Although high correlations among the SRP forms were found, significant differences among the forms were revealed for the three normal subject groups. This suggests that only SRP Forms A and C can confidently be used as alternate forms for measuring the language production of persons with normal language. In order to use the SRP and the %IU/Min metric confidently with normal speakers, further investigation of this metric and its reliability is necessary.

Finally, the results of this investigation indicate that the %IU/Min is a more sensitive metric than the %IU in differentiating individuals with aphasia from those with normal language. However, given that 8% of young normal and 32% old normal subjects were misclassified as having aphasia, the SRP and the %IU/Min metric should be used with some caution for the purposes of detecting aphasia or selecting normal, non-aphasic subjects.

REFERENCES

- Arvedson, C.J., McNeil, M., & West, T. (1985). Prediction of Revised Token Test overall, subtest, and linguistic unit scores by two shortened versions. *Clinical Aphasiology*, *15*, 57–63.
- Balota, D.A., Dolan, P.O., & Duchek, J.M. (2000). Memory changes in healthy older adults. In E. Tulving & F.I.M. Craik (Eds.), *The Oxford handbook of memory* (pp. 395–409). Oxford: Oxford University Press.
- Bayles, K., & Tomoeda, C. (1993). *Arizona Battery for Communication Disorders of Dementia*. Tucson: Canyonlands Publishing.
- Berndt, R.S., Wayland, S., Rochon, E., Saffran, E., & Schwartz, M. (2000). *Quantitative production analysis (QPA)*. Philadelphia: Psychology Press.
- Brodsky, M.B., McNeil, M.R., Park, G.H., Fossett, T.R.D., Timm, N.H., & Doyle, P.J. (2000). *An auditory serial position effect in story retelling for normal adult subjects and for persons with aphasia*. Paper presented at the Academy of Aphasiology, Montreal, Canada.
- Brookshire, R.H., & Nicholas, L.E. (1997). *Discourse comprehension test*. Minneapolis, Minnesota: BRK Publishers.
- Chafe, W.L. (1980). The development of consciousness in the production of a narrative. In R.O. Freedle (Series Ed.) & W.L. Chafe (Vol. Ed.), *Volume III in the series Advances in discourse processes. The Pear stories: Cognitive, cultural, and linguistic aspects of narrative production* (pp. 9–50). Norwood, NJ: Ablex Publishing Corp.
- Carlesimo, G.A., Sabbadini, M., Fadda, L., & Caltagirone, C. (1997). Word-list forgetting in young and elderly subjects: Evidence for age-related decline in transferring information from transitory to permanent memory condition. *Cortex*, *33*, 155–166.
- Doyle, P.J., McNeil, M.R., Park, G.H., Goda, A.J., Spencer, K., Lustig, A., & Carroll, B. (2000). Linguistic validation of four parallel forms of a story retelling procedure. *Aphasiology*, *14*, 537–549.
- Doyle, P.J., McNeil, M.R., Spencer, K.A., Goda, A.J., Cottrell, K., & Lustig, A.P. (1998). The effects of concurrent picture presentations on retelling of orally presented stories by adults with aphasia. *Aphasiology*, *12*, 561–574.
- Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders*. Philadelphia: Lea & Febiger.
- Korsnes, M.S., & Gilinsky, A.S. (1993). Aging and serial list picture memory. *Perceptual and Motor Skills*, *76*, 1011–1014.
- Korsnes, M.S., & Magnussen, S. (1996). Age comparisons of serial position effects in short-term memory. *Acta Psychologica*, *94*, 133–143.
- McNeil, M.R., Doyle, P.J., Fossett, T.R.D., Park, G.H., & Goda, A.J., (2001). Reliability and concurrent validity of the information unit scoring metric for the Story Retell Procedure. *Aphasiology*, *10/11*, 991–1006.
- McNeil, M.R., & Prescott, T. (1978). *The Revised Token Test*. Austin, TX: Pro-Ed.
- McNeil, M.R., Small, S.L., Masterson, R.J., & Fossett, T.R.D. (1995). Behavioral and pharmacological treatment of lexical-semantic deficits in a single patient with primary progressive aphasia. *American Journal of Speech-Language Pathology*, *4*, 76–87.
- Nicholas, L.E., & 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., & 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.
- Oelschlaeger, M.L., & Thorne, J.C. (1999). Application of the correct information unit analysis to the naturally occurring conversation of a person with aphasia. *Journal of Speech Language and Hearing Research*, *42*, 636–648.
- Porch, B. (1981). *Porch Index of Communicative Ability*. Palo Alto, CA: Consulting Psychologists Press.
- Ulatowska, H.K., Chapman, S.B., Highley, A.P., & Prince, J. (1998). Discourse in healthy old-elderly adults: A longitudinal study. *Aphasiology*, *12*, 619–633.
- Ulatowska, H.K., Macaluso-Haynes, S., & North, A.J. (1980). Production of narrative and procedural discourse in aphasia. *Clinical Aphasiology*, *9*, 17–27.
- Wambaugh, J.L., Thompson, C.K., Doyle, P.J., & Camarata, S. (1991). Conversational discourse of aphasic and normal adults: An analysis of communicative functions. *Clinical Aphasiology*, *20*, 343–353.
- 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.