

## Stability of Word Retrieval and Discourse Measures in Aphasia

Several investigators report the effects of naming treatment on discourse production (Antonucci, 2009; Boyle, 2004; Boyle & Coelho, 1995; Conley, Coelho, & Boyle, 2003; Edmonds, Nadeau, & Kiran, 2008; Law, Wong, Sung, & Hon, 2006). However, little is known about the session-to-session stability of word-finding in discourse. Because clinical investigators sometimes claim that changes in the pattern of word-finding errors are a result of treatment, it is important to assess how stable these error patterns are without treatment. The Test of Word Finding in Discourse (TWFD; German, 1991) provides this information for children, but no data about the session-to-session stability of scores is available for aphasic adults. Additionally, a recent report (Cameron, Wambaugh, & Mauszycki, 2009) suggested that individual variability on informativeness and efficiency measures of connected speech (frequently used to assess discourse changes in aphasic adults) was greater than that reported by Nicholas and Brookshire (1993). Additional investigation of the stability of all of these discourse measures is important to provide a dependable basis for investigating changes resulting from treatment.

The purpose of this study was to analyze the session-to-session stability of measures of word-finding behavior in discourse and to re-examine the session-to-session stability of measures of informativeness and efficiency of connected speech.

### Method

#### *Participants*

The participants were 12 right-handed native-English aphasic speakers with anomia as a prominent characteristic in connected speech. None had other history of neurologic impairment and none received concomitant speech-language treatment. Table 1 contains demographic information and test results.

#### *Procedures*

Stimuli to elicit discourse production were Set A from the tasks developed by Nicholas and Brookshire (1993; 1994), consisting of two drawings of complex scenes, one picture sequence that relates a story, one request for personal information, and one request for procedural information. Using the procedures described by Nicholas and Brookshire (1993), discourses were elicited in three sessions separated by 2 to 7 days. All sessions were audiotaped then orthographically transcribed by a graduate student. The author independently checked the transcriptions. Disagreements were resolved prior to scoring.

Transcripts of the discourse tasks were analyzed using the procedures for segmenting, scoring, and for calculating total T-units described by German (1991). Operational definitions of error word-finding behaviors in discourse are in Appendix A. The percentage of T-units containing evidence of any word-finding behavior, as well as the percentage of T-units containing each category of word finding behavior, were calculated. The procedures described by Nicholas and Brookshire (1993) to determine words, correct information units (CIUs), CIUs per minute (CIUs/min), percent of words that were CIUs (%CIUs), and accuracy and completeness of main concepts (Nicholas & Brookshire, 1995) were applied.

After training and practice with the scoring systems, a graduate student scored all transcripts. The author independently scored one of each participant's randomly selected transcripts (one-third of all transcripts). Point-to-point interjudge agreement exceeded 88% for

T-units, word finding behaviors, number of words, and number of CIUs. Interjudge reliability exceeded 80% for each of the four main concept scoring categories.

## Results

### *Word finding analyses*

P3's data was excluded because he produced few T-units in relation to fragments (German, 1991). To assess the extent to which the remaining participants' word finding scores in Session 1 were related to their scores in subsequent sessions, Pearson product-moment correlation coefficients were calculated (Table 2). Correlations for the percentage of T-Units with one or more word-finding behaviors (%TWFB) ranged from .76 to .94, suggesting a strong relationship among session scores. To assess how accurately one could predict an individual's subsequent scores from an earlier score, the standard error of measurement (SEM) was calculated. In general, small SEMs indicate greater session-to-session stability. The SEMs for %TWFB were relatively small in relation to the means for each session. Furthermore, the percentage change (PC) from a session's group mean accounted for by a change in score of 1 SEM between sessions ranged from 4 to 7, indicating acceptable session-to-session stability for this measure.

Results for the percentage of T-Units that contained one or more of each error word-finding behavior were far less favorable. Inspection of standard deviations (Table 2) reveals wide variability. Percentage change measures were unacceptably large even for the few behaviors with strong correlation coefficients.

Each word-finding error category was also examined to see whether it occurred in a similar percentage of T-units from one session to the next. The figure reveals that ten of eleven participants (P12 being the exception) demonstrated session-to-session changes in the distribution of the types of word-finding behaviors. These results will be discussed in detail, but one example demonstrates the importance of this finding: P6 and P9 demonstrated a reduction in phonemic paraphasias and an increase in verbal paraphasias. Clinical investigators of semantically based confrontation naming word retrieval treatments often claim that such a change in error patterns indicates that the participant is better able to access the semantic system after treatment. However, these results demonstrate that at the discourse level such changes in error patterns can be part of the normal variability of the behavior and not due to treatment.

### *Measures of informativeness, efficiency, and accuracy/completeness of main concepts*

Results for measures of informativeness, efficiency, and accuracy/completeness of main concepts (Table 3) include data from all 12 participants. To assess the extent to which participants' scores in one session were related to their scores in subsequent sessions, Pearson product-moment correlation coefficients were calculated. The correlations for the number of CIUs, words per minute, and CIUs per minute were strong (.85 to .99), similar to those reported by Nicholas and Brookshire (1993). The correlations for number of words were strong, but lower than those reported by Nicholas & Brookshire. However, the correlations for %CIUs were only moderate, and were much lower than those of the Nicholas and Brookshire study. The source of the weaker correlations for these two measures was the variability of a single participant, P4, who doubled the number of words he produced from the first to the second session without a concomitant increase in the number of CIUs. P4's aphasia was the most severe of the participants included in this sample, which might account for his more variable performance.

Correlation coefficients for the Set A main concept analyses were similar to those reported by Nicholas & Brookshire (1995) for the full set of 10 stimuli except for the Inaccurate and the combined Accurate Incomplete + Inaccurate categories. The range of Inaccurate responses was extremely limited in this study, accounting for the weak correlations for these categories.

### Discussion

The results suggest that a general measure of word-finding difficulty in discourse, %TWFB, was stable from session to session, whereas measures of individual categories of word-finding behavior were not. For all participants, the proportion of individual categories of word-finding errors contributing to total errors changed from session to session. Clinical researchers should exercise caution in asserting that changes in error type arise from treatment unless they can demonstrate session-to-session stability of the error distributions prior to treatment. Measures of informativeness, efficiency, and main concept production replicated the stability reported by Nicholas & Brookshire (1993, 1995) at the group level, but reinforce findings of variability of some individual participants reported by Cameron, Wambaugh, and Mauszycki (2009), supporting their recommendation that such individual variability be considered by clinicians and researchers when reporting change over time.

## References

- Antonucci, S.M. (2009). Use of semantic feature analysis in group aphasia treatment. *Aphasiology*, 23, 854-866.
- Boyle, M. (2004). Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. *American Journal of Speech Language Pathology*, 13, 236-249.
- Boyle, M., & Coelho, C.A. (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech Language Pathology*, 4, 94-98.
- Brookshire, R.H. & Nicholas, L.E. (1994). Test-retest stability of measures of connected speech in aphasia. *Clinical Aphasiology*, Vol. 22, 119-133.
- Cameron, R., Wambaugh, J., & Mauszycki, S. (2009). Quantifying the informativeness and efficiency of connected speech: An analysis of individual variability over time. Poster presented at Clinical Aphasiology Conference, 5/28/09.
- Conley, A. & Coelho, C.A. (2003). Treatment of word retrieval impairment in chronic Broca's aphasia. *Aphasiology*, 17, 203-211.
- Edmonds, L.A. Nadeau, S.E., Kiran, S. (2008). Effects of Verb Network Strengthening Treatment (VNeST) on lexical retrieval of content words in sentences in persons with aphasia. *Aphasiology*, 23, 402-424.
- German, D.J. (1990). *Test of Adolescent/Adult Word Finding*. Austin, TX: pro-ed.
- German, D.J. (1991). *Test of Word Finding in Discourse (TWFD): Administration, scoring, interpretation, and technical manual*. Austin, TX: pro-ed.
- Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *Boston Naming Test*. Philadelphia: Lea & Febiger.
- Kertesz, A. (1982). *Western Aphasia Battery*. New York: The Psychological Corporation.
- Law, Wong, Sung, & Hon (2006) A study of semantic treatment of three Chinese anomic patients. *Neuropsychological Rehabilitation*, 16, 601-629
- 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.

Table 1. Participants' demographic information and test results.

	<b>P1</b>	<b>P2</b>	<b>P3</b>	<b>P4</b>	<b>P5</b>	<b>P6</b>	<b>P7</b>	<b>P8</b>	<b>P9</b>	<b>P10</b>	<b>P11</b>	<b>P12</b>
Age	57	67	61	61	70	87	65	38	52	50	80	51
Gender	M	F	M	M	M	M	M	F	F	M	M	M
Education (years)	12	12	12	12	18	12	16	12	12	12	19	16
MPO	65	15	59	14	15	36	37	38	64	7	14	72
<i>WAB AQ</i>	82		67	54.5	90.6	72.2	86.6	67.4	70.2	82	61.2	46.3
Aphasia type	BA	BA	BA	BA	AA	AA	AA	AA	CA	CA	WA	WA
<i>TAWF(106)</i>	67	82	68	28	84	63.0	53	56	57	84	28	8
Standard Score	63	88	76	<70	90	72.0	<70	<52	<58	78	<70	<58
Percentile Rank	0.2	19	4	<1	23	2	<1	<0.1	<0.1	6	<1	<0.1
Etiology	LCVA	L CVA	L CVA	L CVA	L CVA	L CVA	L CVA	TBI	L CVA	L CVA	L CVA	L CVA

MPO = months post onset

*WAB AQ* = *Western Aphasia Battery Aphasia Quotient*

BA = Broca's aphasia; AA = anomie aphasia; CA = conduction aphasia; WA = Wernicke's aphasia

*TAWF* = *Test of Adolescent/Adult Word Finding*

Table 2. Average scores of participants on word retrieval measures in each session and Pearson product-moment correlation coefficients (r), standard error of measurement (SEM) values, and the percent of the mean represented by a change of 1 SEM (PC) for differences in participants' performance between Sessions 1 and 2, Sessions 2 and 3, and Sessions 1 and 3 on the measures of word retrieval.

		%TWFB	VP	IS	PP	N	Rep	Ref	E	TF	D	C
Session 1	Mean	76	13.9	27.3	8.6	3.5	40.5	35.6	28.2	17.9	9.2	4.1
	SD	16.78	9.40	19.051	11.23	5.41	15.89	13.1	19.57	29.50	15.69	6.85
	Range	53-100	0-25	3-69	0-30	0-19	8-69	8-57	0-67	0-100	0-54	0-23
Session 2	Mean	72	13.7	28.7	10.4	3.7	41.2	41.7	22.5	17.1	10.1	2.5
	SD	16.52	8.67	26.59	11.23	4.73	19.88	16.23	18.22	26.55	19.73	3.11
	Range	44-100	0-29	0-70	0-33	0-13	14-87	17-70	0-63	0-93	0-67	0-10
Session 3	Mean	74	14.7	25.5	7.2	1.5	38.6	40.8	22.0	14.7	6.5	4.6
	SD	13.53	8.52	23.89	8.30	2.54	19.65	17.06	17.11	25.39	11.61	6.55
	Range	50-94	0-25	0-69	0-21	0-8	21-91	14-72	0-46	0-88	0-38	0-18
Sessions 1 to 2	r	0.90	0.38	0.69	0.68	0.30	0.46	0.60	0.91	0.90	0.95	0.73
	SEM	5	7	11	6	5	12	8	6	9	3	4
	PC	5.31	53.24	38.92	73.42	127.40	28.78	23.07	20.38	52.13	36.36	86.98
Sessions 2 to 3	r	0.94	0.55	0.87	0.75	0.76	0.85	0.53	0.89	0.97	0.94	0.73
	SEM	4	6	9	4	1	8	12	6	5	3	3
	PC	4.04	38.76	34.33	58.37	80.53	19.98	28.63	26.02	31.60	43.65	72.9
Sessions 1 to 3	r	0.75	0.63	0.71	0.78	0.04	0.55	0.66	0.83	0.93	0.97	0.56
	SEM	8	5	14	5	5	13	9	8	7	3	2
	PC	6.63	38.21	50.03	50.49	124.36	32.32	23.19	33.93	40.21	34.20	81.39

Note: %TWFB = percentage of T-Units with one or more word-finding behaviors; VP = verbal paraphasias; IS = initial sounds; PP = phonemic paraphasias; N = neologisms; Rep = repetitions; Ref = reformulations; E = empty words; TF = time fillers; D = delays; C = comments; see Appendix for operational definitions and examples.

Table 3. Pearson product-moment correlation coefficients for differences in participants' performance between Sessions 1 and 2, Sessions 2 and 3, and Sessions 1 and 3 on the measures of informativeness, efficiency, and the accuracy and completeness of main concepts.

	#words	#CIUs	WPM	CIUs/min	%CIUs	AC	AI	IN	AB	AI + IN
Sessions 1 to 2	.84	.85	.99	.97	.61	.85	.70	.34	.83	.42
Sessions 2 to 3	.78	.89	.99	.99	.95	.86	.86	.23	.92	.75
Sessions 1 to 3	.74	.88	.99	.96	.64	.88	.80	.41	.87	.75

Note: CIUs = correct information units, WPM = words per minute, %CIUs = percentage of all words that were CIUs, AC = accurate and complete, AI = accurate but incomplete, IN = inaccurate, AB = absent.

## Appendix

Operational definitions and examples of error word finding behaviors in discourse.

Response	Operational Definition	Examples
Verbal Paraphasia	An unintended substitution of one word for another	
Initial sounds	Partial production of the target or partial production of a substitution	“si”/sink
Phonemic Paraphasia	A nonword obviously related in sound to the target and fluently produced.	“pelim”/penguin
Neologism	A nonword with no, or only a remote, relation to the target.	“budisky”/spider
Repetition	Any word inappropriately uttered twice in a T-Unit	
Reformulation	Changes or modifications to one or more previous words in a T-Unit	“Well I go to the I usually watch programs.”
Empty/Indefinite words	A vague or indefinite word applied to the target	“That’s very good stuff”
Time fillers	Verbalizations of vowel sounds or syllables produced to maintain the listener’s attention during word-retrieval difficulty. Mark only when <i>three or more</i> occur in a T-Unit.	
Delays	Any prolonged pause of 6 seconds or more with or without time fillers within a T-unit.	
Comment	A comment on the task or the language process.	“The man is I can’t tell you his name.”



## Figure Caption

Individual participant data for each word-finding category across the three sessions reported as the percentage of T-units in which the behavior occurred. P1-B1 = Participant 1, Baseline session 1; P1-B2 = Participant 1, Baseline session 2, etc. Data for Participant 3 was not used in these analyses; see text.







