Journal of Speech and Hearing Research, Volume 36, 338–350, April 1993

A System for Quantifying the Informativeness and Efficiency of the Connected Speech of Adults With Aphasia

Linda E. Nicholas
Robert H. Brookshire
Department of Veterans Affairs
Medical Center
and
University of Minnesota
Minneapolis

A standardized rule-based scoring system, the Correct Information Unit (CIU) analysis, was used to evaluate the informativeness and efficiency of the connected speech of 20 non-brain-damaged adults and 20 adults with aphasia in response to 10 elicitation stimuli. The interjudge reliability of the scoring system proved to be high, as did the session-to-session stability of performance on measures. There was a significant difference between the non-brain-damaged and aphasic speakers on each of the five measures derived from CIU and word counts. However, the three calculated measures (words per minute, percent CIUs, and CIUs per minute) more dependably separated aphasic from non-brain-damaged speakers on an individual basis than the two counts (number of words and number of CIUs).

KEY WORDS: aphasia, speech, informativeness, efficiency

Analyses of the connected speech of adults with aphasia have focused primarily on how their speech conforms to standard language rules and patterns rather than how well it communicates information to listeners. Clinicians and investigators who wish to quantify changes in the informativeness of the connected speech of adults with aphasia in response to treatment or in response to manipulation of experimental variables have been hampered by the scarcity of standard measures for characterizing this aspect of connected speech. Yorkston and Beukelman (1980) have provided the only published system for quantifying the amount of information conveyed by aphasic speakers. They had 78 non-brain-damaged adults describe the "cookie theft" picture from the Boston Diagnostic Aphasia Examination (BDAE) (Goodglass & Kaplan, 1983). From these descriptions, they developed a list of what they called "content units." They defined a content unit as "a grouping of information that was always expressed as a unit" and was mentioned by at least 1 of their 78 non-brain-damaged subjects. They reported that content units per minute differentiated the speech of aphasic subjects from that of non-brain-damaged subjects. They also demonstrated, based on the performance of 1 aphasic subject, that two measures—number of content units and content units per minute—were potentially sensitive measures of change in connected speech as a consequence of treatment. Others have modified or expanded upon Yorkston and Beukelman's system (Golper, Thorpe, Tompkins, Marshall, & Rau, 1980; Shewan, 1988). However, Yorkston and Beukelman's system and its variants have a major limitation. Because their content units are specific to a single elicitation picture, their system can be used only with that elicitation stimulus unless content units are defined for other stimuli.

In response to this problem, others have devised rule-based systems for quantifying the amount of information conveyed by speakers for use in their own research

0022-4685/93/3602-0338$01.00/0

(Bernstein-Ellis, Wertz, & Shubitowski, 1987; Busch, Brookshire, & Nicholas, 1988; Gaddie, Kearns, & Yedor, 1991; Gleason et al., 1980; Kearns, 1985). During our own work with a rule-based system for scoring the informativeness of a speaker's words, we have found that such a scoring system can be used to reliably quantify speech elicited with a variety of stimuli, ranging from single pictures and picture sequences (Brenneise-Sarshad, Nicholas, & Brookshire, 1991; Correia, Brookshire, & Nicholas, 1990; Potechin, Nicholas, & Brookshire, 1987) to requests for personal and procedural information (MacLennan, Nicholas, Morley, & Brookshire, 1991; Schumacher & Nicholas, 1991).

At this time, no standardized rule-based system for quantifying the informativeness of connected speech elicited with a variety of stimuli has been published. In this paper we will (a) describe such a system (Correct Information Unit [CIU] analysis), report the reliability of the CIU analysis, report the session-to-session stability of the CIU measures, and report the sensitivity of CIU measures for discriminating the connected speech of aphasic speakers from that of non-brain-damaged speakers.

Method

Subjects

Subjects were 20 non-brain-damaged adults (10 male, 10 female) and 20 adults with aphasia (18 male, 2 female). All were right-handed and native speakers of English. Aphasic subjects were recruited from past and current caseloads of speech-language pathology clinics in the Minneapolis–St. Paul metropolitan area. Each was at least 3 months post onset of a single left-hemisphere thromboembolic brain injury. They were diagnosed as aphasic by speech-language pathologists based on results of standard tests. The aphasic subjects demonstrated a range of type and severity of aphasia. Six exhibited nonfluent (essentially Broca's) aphasia and 14 exhibited fluent aphasia (fluent speech with literal paraphasias and word retrieval difficulty). Type of aphasia (fluent or nonfluent) was determined by the two authors from audiotaped samples of conversational speech and picture description and from performance on the auditory comprehension subtests of the BDAE. If the two judges disagreed on a subject's aphasia type, a third judge rated it. Two of the three judges had to agree on a subject's aphasia type for the subject to be included in the study. The severity of their aphasia was estimated by their overall percentile on a four-subtest shortened version (SPICA) (Disimoni, Keith, & Darley, 1980) of the Porch Index of Communicative Ability (Porch, 1971). Aphasic subjects also were tested with the Boston Naming Test (BNT) (Kaplan, Goodglass, & Weintraub, 1983). Non-brain-damaged subjects were nonhospitalized and noninstitutionalized adults who reported no history of neurologic or psychiatric problems. They ranged in age from 50 to 73 years (M = 64.2; SD = 7.0) and in years of education from 8 to 16 (M = 12.8; SD = 2.2). Age and education were similar for the non-brain-damaged and aphasic subjects. Descriptive information and test scores for aphasic subjects are presented in Table 1.

TABLE 1. Descriptive information for 20 aphasic subjects.

	Age	Education	Months PO	SPICA %ile	BNT
MEAN	64.9	13.1	56.2	63.7	31.0
SD	6.8	1.7	62.7	14.5	12.1
RANGE	51–77	10–16	3–192	40–85	0–48

Note. Months PO = months postonset of aphasia. SPICA %ile = overall percentile on a four-subtest version of the Porch Index of Communicative Ability. BNT = number correct of 60 possible on the Boston Naming Test.

Stimulus Materials

Connected speech was elicited with 10 stimuli, all of which were likely to be familiar to most North American adults. Four were single pictures, two were picture sequences, two were requests for personal information, and two were requests for procedural information. Two of the single pictures were from standard aphasia tests (the "cookie theft" picture from the Boston Diagnostic Aphasia Examination (BDAE) (Goodglass & Kaplan, 1983) and the "picnic" picture from the Western Aphasia Battery (WAB) (Kertesz, 1982). The other two single pictures and the two picture sequences were drawn to the authors' specifications by a professional artist. They are shown in Appendix A. The two single pictures each depicted a story-like situation with a central focus and interactions among pictured elements. Each implied a series of events leading up to the pictured scene, and each suggested events that were likely to follow the pictured situation. The two-picture sequences each contained six pictures that related a story. The requests for personal information (Tell me what you usually do on Sundays. Tell me where you live and describe it to me) and the requests for procedural information (Tell me how you would go about doing dishes by hand. Tell me how you would go about writing and sending a letter) were selected from larger sets of requests based on the results of a pilot study, which suggested that they would elicit speech samples with a reasonable amount and consistency of content across speakers.

Procedures

All testing sessions were conducted in a quiet room, free from distractions. The subject and the examiner sat side-by-side at a table that held an audiocassette recorder and a microphone. Each subject's spoken responses to the elicitation stimuli, together with any prompts delivered by the examiner, were recorded on audiotape.

Pretest and practice. To determine whether subjects had adequate vision for the task, each subject was asked to match line drawings of single objects to the same objects in a composite line drawing (the speech elicitation picture from the Minnesota Test for Differential Diagnosis of Aphasia, Schuell, 1972). Subjects were shown, one at a time, six single objects from the composite picture and asked to point to the same object in the composite picture. Subjects had to identify correctly all six to participate in the study. No subject failed to pass this visual screening test.

To determine if subjects' speech was sufficiently intelligible to permit accurate transcription and if subjects produced enough speech to permit meaningful scoring of content, each subject was asked to describe a three-picture sequence that related a short story. To be included in the study, a subject had to produce at least 10 intelligible, relevant, nonrepeated words in response to the picture sequence. If a subject failed to meet the criterion for inclusion on a first try, the subject was given a second (and final) chance to describe the same picture sequence. Two subjects who exhibited severe non-fluent aphasia were unable to meet this criterion and were excluded from the study.

The 40 subjects who qualified for the study were given a short interval of practice and training with two stimuli that were not used in the study (Tell me what you like to spend your time doing. Tell me how you would go about making a sandwich). Instruction and feedback were provided, as needed, until the examiner felt that the subject understood the task. Feedback and instruction generally took the form of letting the subject know if his or her response was considered satisfactory in length and content. If it was not considered satisfactory, the subject was given suggestions about other information that could have been provided.

Elicitation of speech samples. The stimulus pictures and spoken requests were presented individually to subjects in random order. When the eliciting stimulus was a picture or picture sequence, it was placed on the table in front of the subject when the examiner asked the subject to talk about it, and was left there until the subject finished. Subjects were instructed to tell what they saw happening in the picture(s). When the eliciting stimulus was a spoken request for personal or procedural information, the examiner placed a card on which the request was typed in bold quarter-inch letters on the table in front of the subject, and left it there until the subject finished. Subjects were asked to try to talk about each stimulus for about 1 minute. If a subject stopped talking before producing at least 15 sec of speech, he or she was prompted once with "Can you tell me more?" No further prompts were given. The examiner provided no feedback regarding the accuracy or appropriateness of subjects' responses, but did provide occasional social continuants such as "uh-huh" and head nods.

A different random order of stimuli was established for each subject, and each subject responded to the 10 stimuli in the same order in three sessions. (Subjects were told at the beginning of the first session that they would be asked to talk about the stimuli three times.) The first two sessions took place on the same day and were separated by a 10-min break. The third session took place 7 to 10 days following the first two sessions. Practice was provided only at the beginning of the first session.

Transcribing speech samples. Subjects' recorded speech samples were orthographically transcribed by a speech-language pathologist familiar with the speech of aphasic adults. A second speech-language pathologist independently checked the transcriptions against the tapes. Transcription disagreements between the two were resolved by the first author.

Scoring and timing speech samples. The rules used to score words and correct information units (CIUs) are provided in Appendix B.[1] To be included in the word count, words had to be intelligible in context but did not have to be accurate, relevant, or informative relative to the eliciting stimulus. To be included in the CIU count, words had to be accurate, relevant, and informative relative to the eliciting stimulus. Words did not have to be used in a grammatically accurate manner to be counted as CIUs. Each CIU consisted of a single word, and only words that were included in the word count could be counted as correct information units.

The 1,200 speech samples (40 subjects × 10 speech samples × 3 sessions) were scored by the first author and another speech-language pathologist for number of words and CIUs. To ensure that the two scorers were consistent in their interpretation of the scoring rules, they first independently scored transcripts for 6 subjects responding to the 10 eliciting stimuli. Then they compared their scoring, discussed disagreements, and clarified misunderstandings. Following this practice scoring, they each scored approximately half of the remaining transcripts. (Interjudge reliability data are presented in the Results section of this report.)

After words and CIUs were identified in each transcript, the speech samples were timed. Time occupied by examiner prompts and by patient commentary that preceded or followed their response to the eliciting stimulus was subtracted from the overall time for each sample.

Time and word and CIU counts were used to calculate three measures: (a) words per minute (WPM), (b) percent of words that were correct information units (%CIUs), and (c) correct information units per minute (CIUs/min).

Results

Interjudge Reliability of Scoring

To assess interjudge reliability, the two scorers who scored the transcripts for this study both independently scored a systematically selected, representative sample of the transcripts (10 speech samples for each of 6 non-brain-damaged and 6 aphasic subjects). The aphasic subjects were representative of the group in severity and type of aphasia. Four of the subjects exhibited fluent aphasia and 2 exhibited nonfluent aphasia. SPICA overall percentiles for these 6 subjects ranged from 45 to 82 ($M = 65.5$; $SD = 12.8$). Point-to-point interjudge percent agreement for number of words and number of CIUs was calculated with the following formula:

$$[\text{total agreements}/(\text{total agreements} + \text{total disagreements})] \times 100.$$

Interjudge reliability exceeded 98% for words and 90% for CIUs for all 12 subjects and did not appear to be strongly

[1]The scoring rules in Appendix B are the product of numerous modifications over the course of six studies to improve their clarity, to eliminate ambiguity, and to correct deficiencies. In some cases, theoretical and linguistic considerations had to be tempered by practical factors such as ease and consistency of scoring. Those who use the scoring rules should know that the amount of explanation provided relative to scoring a given aspect of the speech samples may reflect more the difficulty in achieving reliable scoring for that aspect than its theoretical or clinical importance.

TABLE 2. Mean absolute differences, 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 subjects' performance between Sessions 1 and 2 and Sessions 1 and 3.

		#words	#CIUs	WPM	%CIUs	CIUs/ min
Non-brain-damaged (*n* = 20)						
Sessions 1 to 2	*M*	13	11	7	2	9
	SD	13	11	6	1	6
	Range	1–42	2–38	0–22	0–4	2–24
	r	.93	.93	.92	.94	.88
	SEM	9	8	6	1	7
	PC	8.0	8.3	3.6	1.2	4.9
Sessions 1 to 3	*M*	14	11	9	2	8
	SD	11	9	7	1	7
	Range	1–42	0–30	0–27	0–5	1–24
	r	.95	.96	.90	.96	.90
	SEM	8	7	7	1	6
	PC	7.1	7.2	4.2	1.2	4.2
Aphasic (*n* = 20)						
Sessions 1 to 2	*M*	9	5	7	3	5
	SD	9	5	5	1	5
	Range	0–40	0–19	0–21	1–5	0–15
	r	.94	.97	.98	.98	.98
	SEM	8	4	6	2	4
	PC	10.3	8.3	7.4	3.2	8.2
Sessions 1 to 3	*M*	11	7	6	3	4
	SD	10	7	5	2	4
	Range	0–43	0–28	0–19	0–6	0–15
	r	.89	.91	.98	.98	.97
	SEM	11	7	5	2	4
	PC	14.1	14.6	6.2	3.2	8.2

Note. Difference scores are for average performance across 10 stimuli. CIUs = correct information units. WPM = words per minute. PC = percentage change. Range values are for the range of average change scores across subjects.

affected by length of speech samples or presence, severity, or type of aphasia. Intrajudge reliability was assessed on the first author's scoring of the transcripts of the same 6 aphasic subjects included in the interjudge reliability assessment. Intrajudge reliability exceeded 99% for words and 95% for CIUs for all 6 aphasic subjects.

Session-to-Session Stability of Measures

The session-to-session stability of measures was evaluated in two ways. To provide a robust estimate of the effects of practice with the stimulus materials on subjects' speech, the change in performance between Session 1 and Session 2, which took place on the same day, was measured. The change in performance between Session 1 and Session 3, which were separated by 7 to 10 days, also was measured. This latter difference provided an estimate of session-to-session stability over a time interval that resembles typical clinical practice, in which a patient's performance may be reassessed with the same materials after several days or weeks.

Absolute difference scores were calculated for each subject (negative values were converted to positive values). Absolute difference scores, rather than signed scores, were used so that negative differences would not cancel out positive differences when group statistics were calculated. Furthermore, the amount of change seemed more important to the issue of stability than its direction. The mean absolute difference scores for the 20 non-brain-damaged and 20 aphasic subjects on the five measures are presented in Table 2. Changes in scores between Sessions 1 and 2 were similar to those between Sessions 1 and 3 for all measures. In general, changes in scores between sessions were smaller for the three calculated measures (WPM, CIUs/min, and %CIUs) than for the two counts (number of words and number of CIUs). Changes for both groups were smallest for percent CIUs.

To estimate the extent to which subjects' scores in Session 1 were related to their scores in Sessions 2 and 3, Pearson product-moment correlation coefficients were calculated (Table 2). Correlations ranged from .88 to .98, suggesting that both non-brain-damaged and aphasic subjects' scores in Session 1 were strongly related to their scores in Sessions 2 and 3.

Although the correlation coefficients shown in Table 2 suggest a strong relationship between Session 1 scores and Session 2 and 3 scores, they do not indicate how accurately one could predict an individual's Session 2 or 3 scores from his or her Session 1 score. To provide this information, we calculated the standard error of measurement (SEM) for each of the five measures using the formula

TABLE 3. Average scores of non-brain-damaged and aphasic subjects for 10 elicitation stimuli per session.

		#words	#CIUs	time(sec)	WPM	%CIUs	CIUs/min
Non-brain-damaged (n = 20)							
Session 1	*M*	113	97	41	166	86	143
	SD	36	31	14	22	6	19
	Range	62–176	56–157	19–73	105–198	72–93	92–175
Session 2	*M*	104	89	37	170	87	147
	SD	29	25	12	22	6	21
	Range	59–176	53–142	22–62	119–201	71–93	102–170
Session 3	*M*	101	87	37	167	87	145
	SD	26	24	12	22	6	23
	Range	61–148	54–137	18–69	105–202	75–93	94–178
Aphasic (n = 20)							
Session 1	*M*	78	48	67	81	63	49
	SD	33	24	31	39	15	25
	Range	30–160	15–96	28–156	12–145	23–87	10–102
Session 2	*M*	73	46	61	84	63	52
	SD	28	21	33	39	15	27
	Range	32–120	12–80	30–167	13–149	22–84	11–109
Session 3	*M*	77	48	63	82	63	49
	SD	30	21	26	37	14	24
	Range	37–119	16–86	33–149	15–150	22–85	13–105

Note. Session scores represent average performance across 10 elicitation stimuli. CIUs = correct information units. WPM = words per minute.

$$SEM = SD\sqrt{1 - r},$$

where *SD* is the standard deviation for the obtained score distribution and *r* is the correlation coefficient. The SEM permits one to estimate the consistency (or reliability) with which a test measures performance on repeated test occasions. The chances are about 68 in 100 that a predicted score will not differ from an obtained score by more than ± 1 SEM, and about 95 in 100 that a predicted score will not differ from an obtained score by more than ± 2 SEM. In general, the smaller the SEM, the greater the session-to-session stability of a score. The SEMs for the measures evaluated in this study are provided in Table 2. For both groups, SEM values for Sessions 1 and 2 were similar to those for Sessions 1 and 3 for all measures. The SEM values for the three calculated measures (WPM, %CIUs, and CIUs/min) tended to be smaller than the SEM values for the two count measures (number of words, number of CIUs). The SEM value for percent CIUs for both groups was consistently smaller than that for any other measure. SEM values, by themselves, do not provide a straightforward measure of session-to-session stability, because the acceptability of a particular SEM value depends on the magnitude of the scores that can be expected. For example, an SEM of 10 might be acceptable if scores ranged from 100 to 500, but would not be if scores ranged from 0 to 20. To provide a measure of SEM that takes into account the magnitude of scores on which SEMs are based, we calculated a percentage change (PC) measure, using the formula

$$PC = (SEM/Mean) \times 100,$$

where PC represents the percentage change from the Session 1 group mean accounted for by a change in score of 1 SEM between sessions, and Mean represents the Session 1 group mean for a measure. (The Session 1 group means are given in Table 3.) The results of these calculations are shown in Table 2, and confirm that the three calculated measures yielded greater session-to-session stability than the two count measures. For non-brain-damaged subjects, the 1 SEM percent change values for the two count measures ranged from 7.1 to 8.3, whereas the values for the three calculated measures ranged from 1.2 to 4.9. For aphasic subjects, the 1 SEM percent-change values for the two count measures ranged from 8.3 to 14.6, whereas the values for the three calculated measures ranged from 3.2 to 8.2. For both groups the 1 SEM percent-change scores for percent CIUs were considerably smaller than those for any other measure. There are no rules for determining what constitutes an acceptable SEM. How much instability in scores from session to session is acceptable depends both on the magnitude of the scores and the user's purposes. Therefore, we make no general claims about the "acceptability" of the SEMs reported herein.

Effects of Aphasia on Performance

The non-brain-damaged and aphasic subjects' average scores for number of words, number of correct information units (CIUs), words per minute, percent CIUs, and CIUs per minute for each session are presented in Table 3. On the average, non-brain-damaged subjects produced more words, more CIUs, a higher percentage of CIUs, and more words per minute and CIUs per minute than aphasic subjects. To determine which of the Session 1 differences between the groups were significant, a repeated-measures analysis of variance was calculated for each measure. The experimentwise error rate was adjusted for multiple compar-

TABLE 4. Number of average session scores for 20 aphasic (APH) subjects that were in the range for non-brain-damaged (NBD) subjects' average session scores and that were at or above the cutoff for NBD performance.

Measure	Session	NBD range	#APH in NBD range	NBD cutoff	#APH at or above cutoff	#NBD below cutoff
#words	1	62–176	13	54	14	0
	2	59–176	12	55	13	0
	3	61–148	12	58	12	0
#CIUs	1	56–157	8	47	8	0
	2	53–143	7	49	7	0
	3	54–137	8	49	8	0
WPM	1	105–198	7	131	2	1
	2	119–201	5	134	2	1
	3	105–202	7	125	3	1
%CIUs	1	72–93	6	76	4	2
	2	71–93	7	77	3	2
	3	75–93	5	78	3	2
CIUs/min	1	92–175	2	111	0	2
	2	102–170	2	113	0	2
	3	94–178	1	107	0	2
WPM and	1	see above	2	see above	0	0
%CIUs	2	see above	3	see above	1	0
combined	3	see above	2	see above	0	0

Note. Session scores represent average performance across 10 elicitation stimuli. CIUs = correct information units. WPM = words per minute.

isons by setting the Type 1 (alpha) error rate for each test of significance at $p \leq .01$. This adjustment kept the experimentwise error rate at $p \leq .05$. The non-brain-damaged group produced significantly more words ($F = 10.12$), more CIUs ($F = 29.28$), a greater percentage of CIUs ($F = 40.05$), more words per minute ($F = 71.37$), and more CIUs per minute ($F = 168.04$) than the aphasic group (degrees of freedom = 1,38 and $p \leq .01$ for all comparisons). We did not calculate similar statistics for Sessions 2 and 3. It seemed clear from visual inspection of the data that differences found for Session 1 would hold for Sessions 2 and 3 and carrying out the additional tests would have greatly inflated the experimentwise Type 1 error rate.

Although there were significant differences between the groups on the five measures, there was some overlap in scores between individuals in the two groups for all measures. Table 4 summarizes the nature of the overlap. The amount of overlap was greatest for number of words, with 12 or 13 of 20 aphasic subjects' scores falling within the non-brain-damaged group's range; and the overlap was least for CIUs per minute, with only 1 or 2 of 20 aphasic subjects falling within the non-brain-damaged group's range.

Cutoff scores for "normal" performance were calculated by establishing as the lower limit of non-brain-damaged performance two standard deviations below the non-brain-damaged subjects' group mean for each measure in each session. These cutoff scores are presented in Table 4. Neither count measure dependably separated aphasic from non-brain-damaged subjects. About two-thirds of aphasic subjects' scores fell at or above the cutoff for normal performance for number of words, and slightly less than half fell at or above the cutoff for number of CIUs. The three calculated measures more dependably separated aphasic from non-brain-damaged speakers than the two counts. No aphasic subject's score was at or above the cutoff for CIUs per minute.

Discussion

These results suggest that the informativeness of the connected speech of adults with aphasia can be reliably scored using the procedures described herein. This is true despite the complex and variable nature of the connected speech elicited by our range of stimuli, and regardless of the type and severity of aphasia exhibited by the speakers in our reliability sample. Subjectively, the copious connected speech of our most impaired fluent aphasic subjects was the most challenging to score, but reliability still exceeded 90% for 1 such individual who was included in the reliability sample. We feel that the CIU analysis is likely to provide the most reliable and useful information when it is used to analyze the performance of speakers whose aphasia severity is within the range of the subjects in this study. We do not know whether it will prove reliable or provide clinically useful information if it is applied to the connected speech of adults with more severe aphasia.

In general, the measures were reasonably stable from session to session, whether the sessions were separated by a 10-min break or by a 7- to 10-day period. The three calculated measures (WMP, CIUs/min, and %CIUs) were generally more stable from session to session than the two count measures (number of words, number of CIUs), with percent CIUs demonstrating the greatest stability. We were somewhat surprised that speakers did not exhibit greater changes on the five measures from Session 1 to Sessions 2 and 3. We had expected that Session 1 performance might not be representative of performance in subsequent sessions, because of the practice and familiarity with the stimuli gained in the first session. For this reason, we thought it likely that those who wish to measure changes in aphasic speakers' connected speech over time might want to disregard the results from the first session when establishing baseline performance. This does not appear to be the case. Our results suggest that, for most subjects, performance in the first session is representative of performance in subsequent sessions, if subjects are given prior practice and training with stimuli that are not used in the subsequent sessions. For example, on the percent-CIUs measure, 75% of aphasic subjects' scores and 95% of non-brain-damaged subjects' scores changed by 3% or less. On the CIUs-per-min measure, 90% of aphasic subjects' scores and 75% of non-brain-damaged subjects' scores changed by 10 CIUs per min or less. On the words-per-min measure, 85% of aphasic subjects' scores changed by 10 words per min or less. Performance across non-brain-damaged subjects was somewhat more variable on this measure, with 60% of non-brain-damaged subjects' scores changing by 10 words per min or less and 75% of these subjects' scores changing by 15 words per min or less. These change scores are for differences between Sessions 1 and 3. However, similar results were obtained for differences between Sessions 1 and 2.

In general, these results suggest that the calculated measures, particularly percent CIUs, may yield stable baseline performance against which changes in connected speech with treatment or manipulation of experimental variables can be measured. However, even though the subjects in this study represented a range of type and severity of aphasia, one cannot assume that every aphasic speaker's performance will be stable over repeated assessments. Those who wish to assess changes in an individual's connected speech over time still must establish the stability of the measures used across several baseline sessions.

Although there was a statistically significant difference between the non-brain-damaged and aphasic groups on each of the five measures, the three calculated measures discriminated non-brain-damaged from aphasic speakers better than the two count measures did. The most dependable measure for discriminating the connected speech of non-brain-damaged adults from that of aphasic adults appears to be CIUs per min. No aphasic speaker's performance was at or above the non-brain-damaged cutoff for this measure. Although CIUs per min dependably separated aphasic from non-brain-damaged speakers, this measure is not as informative about performance as joint consideration of words per min and percent CIUs. These two measures, when evaluated concurrently, specify the relative contributions of rate and informativeness to performance. Only 1 aphasic subject performed at or above the non-brain-damaged cutoffs for *both* words per min and percent CIUs and this occurred in only one session.

The three calculated measures provide information about the efficiency of connected speech, although neither percent-CIUs nor words-per-min alone will provide a complete picture of a speaker's efficiency. For example, if two speakers each produced 80% CIUs but one spoke at 140 words per min and the other at 20 words per min, the first would be considered a more efficient speaker than the second. Conversely, if two speakers each spoke at 140 words per min but one produced 80% CIUs and the other produced only 20% CIUs, the first would be considered a more efficient speaker than the second.

The results presented here are based on connected speech samples elicited with 10 stimuli. Preliminary data analyses suggest that connected speech samples elicited with subsets of five of these stimuli yield measures with nearly equivalent stability for many adults with aphasia. (Brookshire & Nicholas, in press, a). We currently are analyzing these data in more depth. This seems a clinically important question, given the typical time constraints imposed on clinicians and the substantial amount of time that often is required to transcribe and score speech samples.

Another area we have begun to evaluate is how well the elicited speech samples in this study represent connected speech in daily life. Our first step was to determine if the word choice pattern for the elicited speech samples was comparable to that of adult-to-adult conversations, as determined by Hayes (1988, 1989). We found that the word choice pattern of our 20 non-brain-damaged speakers was equivalent to that of Hayes's adult-to-adult conversations, except for our subjects' slightly greater reliance on the five most common word types, especially the word *and.* The pattern of word choice for our aphasic speakers generally resembled that of our non-brain-damaged speakers and the adult-to-adult conversations analyzed by Hayes (Brookshire & Nicholas, 1993). We are currently evaluating the relationship between performance on connected speech measures and listeners' judgments of the communicative informativeness and efficiency of the speakers' elicited samples. This will provide information about the relationship between objective measures of elicited speech and subjective impressions of communicative adequacy.

As we worked with the CIU measures in various studies, we realized that there were still important aspects of the informativeness of connected speech that we had not captured. With the CIU analysis, the accuracy, relevance, and informativeness of the words produced by an individual are evaluated but there is no indication of the relative importance of the information that is conveyed or whether important information has been left out. We feel that it is important to know how many of the main concepts (or how much of the gist) a speaker conveys about a picture or topic. A main concept consists of a series of words that expresses an essential piece of information. The set of main concepts for a stimulus provides an outline of the essential information about a picture or topic. We have developed a scoring system to identify the presence, accuracy, and completeness of main concepts in connected speech elicited by the stimuli described in this study (Nicholas & Brookshire, 1993). If information about main concept production is combined with information about the efficiency with which a speaker produces informative words, a more complete 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 and clinically useful analysis of changes in connected speech with language recovery.

Acknowledgments

This research was supported by the Department of Veterans Affairs Medical Research Service. We want to thank Donald MacLennan, Julia Edgar, and Annalisa Margheri for their assistance with this research and Cindy Busch, whose dissertation research stimulated our interest in rule-based systems for scoring the informativeness and efficiency of connected speech. We also want to thank Connie Tompkins, Terry Wertz, and an anonymous reviewer for their editorial assistance.

References

Bernstein-Ellis, E., Wertz, R. T., & Shubitowski, Y. (1987). More pace, less filler: A verbal strategy for a high-level aphasic patient. *Clinical Aphasiology: Conference Proceedings, 17,* 12–22.

Brenneise-Sarshad, R., Nicholas, L. E., & Brookshire, R. (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,* 168–176.

Brookshire, R. H., & Nicholas, L. E. (in press). Test-retest stability of measures of connected speech in aphasia. *Clinical Aphasiology, 22.*

Brookshire, R. H., & Nicholas, L. E. (1993). Word choice in the connected speech of aphasic and non-brain-damaged speakers. *Clinical Aphasiology, 21,* 101–111.

Busch, C. R., Brookshire, R. H., & Nicholas, L. E. (1988). Referential communication by aphasic and nonaphasic adults. *Journal of Speech and Hearing Disorders, 53,* 475–482.

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

Gaddie, A., Kearns, K. P., & Yedor, K. (1991). A qualitative analysis of response elaboration training effects. *Clinical Aphasiology, 19,* 171–183.

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.

Golper, L. A. C., Thorpe, P., Tompkins, C., Marshall, R. C., & Rau, M. T. (1980). Connected language sampling: An expanded index of aphasic language behavior. *Clinical Aphasiology: Conference Proceedings, 10,* 174–186.

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

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

Hayes, D. P. (1989). *Guide to the lexical analysis of texts.* Ithaca, NY: Cornell University Dept of Sociology, Technical Report Series 89–6.

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

Kearns, K. P. (1985). Response elaboration training for patient initiated utterances. *Clinical Aphasiology: Conference Proceedings, 15,* 196–204.

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

MacLennan, D. M., Nicholas, L. E., Morley, G. K., & Brookshire, R. H. (1991). The effects of bromocriptine on speech and language function in a man with transcortical motor aphasia. *Clinical Aphasiology, 20,* 145–156.

Nicholas, L. E., & Brookshire, R. H. (1993). A system for scoring main concepts in the discourse of non-brain-damaged and aphasic speakers. *Clinical Aphasiology, 21,* 87–99.

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

Potechin, G. C., Nicholas, L. E., & Brookshire, R. H. (1987). Effects of picture stimuli on discourse production by aphasic speakers. *Clinical Aphasiology: Conference Proceedings, 17,* 216–220.

Schuell, H. (1972). *The Minnesota Test for Differential Diagnosis of Aphasia.* Minneapolis: University of Minnesota Press.

Schumacher, J. G., & Nicholas, L. E. (1991). Conducting research in a clinical setting against all odds: Unusual treatment of fluent aphasia. *Clinical Aphasiology, 19,* 267–278.

Shewan, C. M. (1988). The Shewan Spontaneous Language Analysis (SSLA) System for aphasic adults: Description, reliability, and validity. *Journal of Communication Disorders, 21,* 103–138.

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.

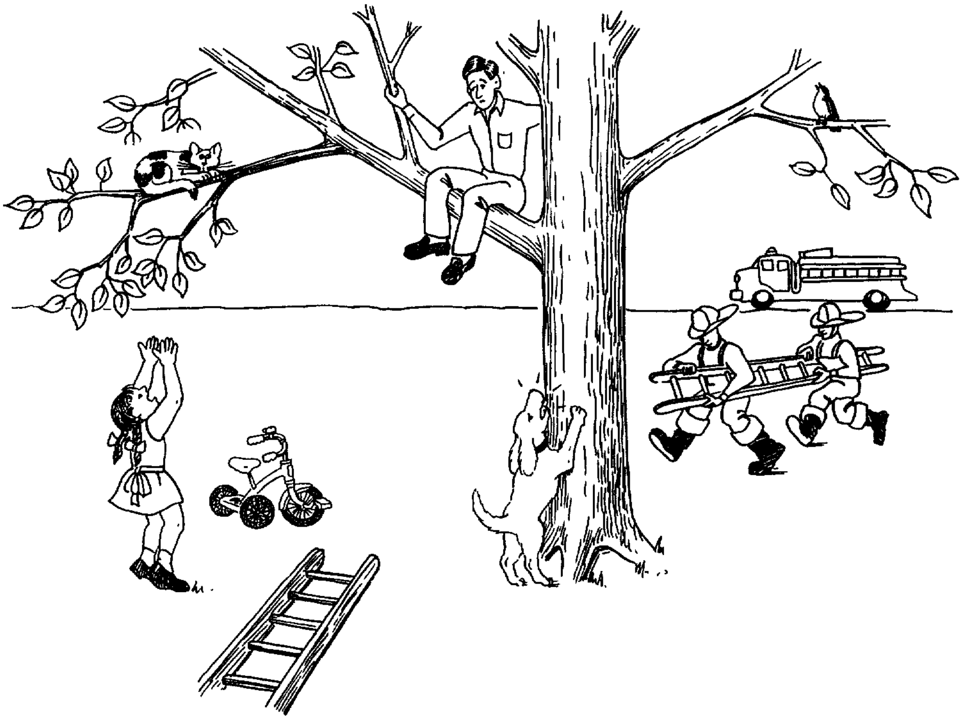
Received June 19, 1992
Accepted October 27, 1992

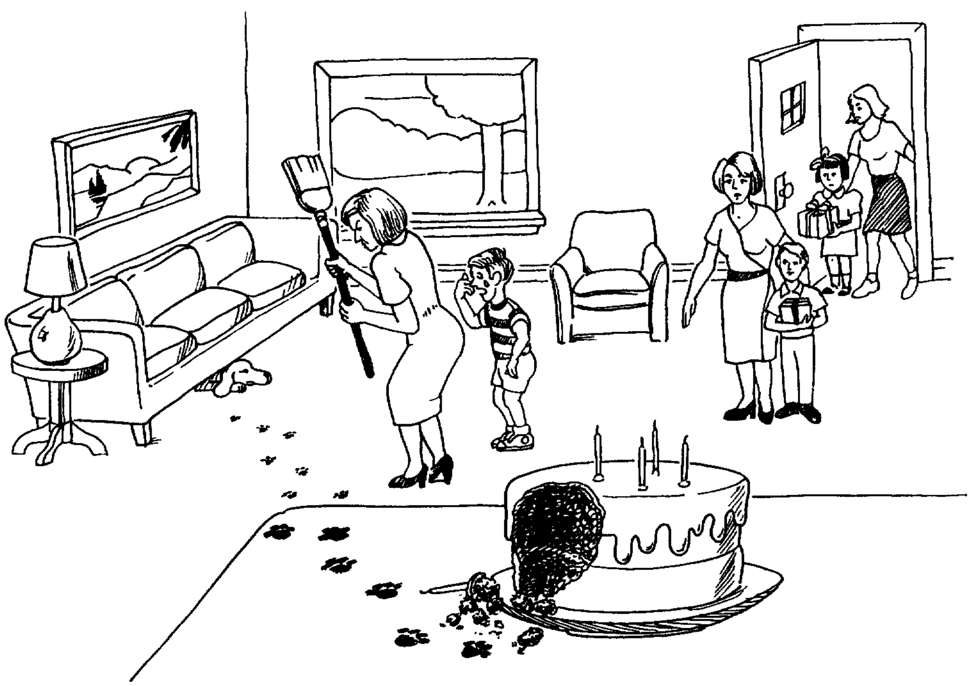
Contact author: Linda E. Nicholas, MA, Aphasia Research (127-A), VA Medical Center, One Veterans Drive, Minneapolis, MN 55417. (The scoring and counting rules in Appendix B are available from the contact author.)

 36 338–350 April 1993

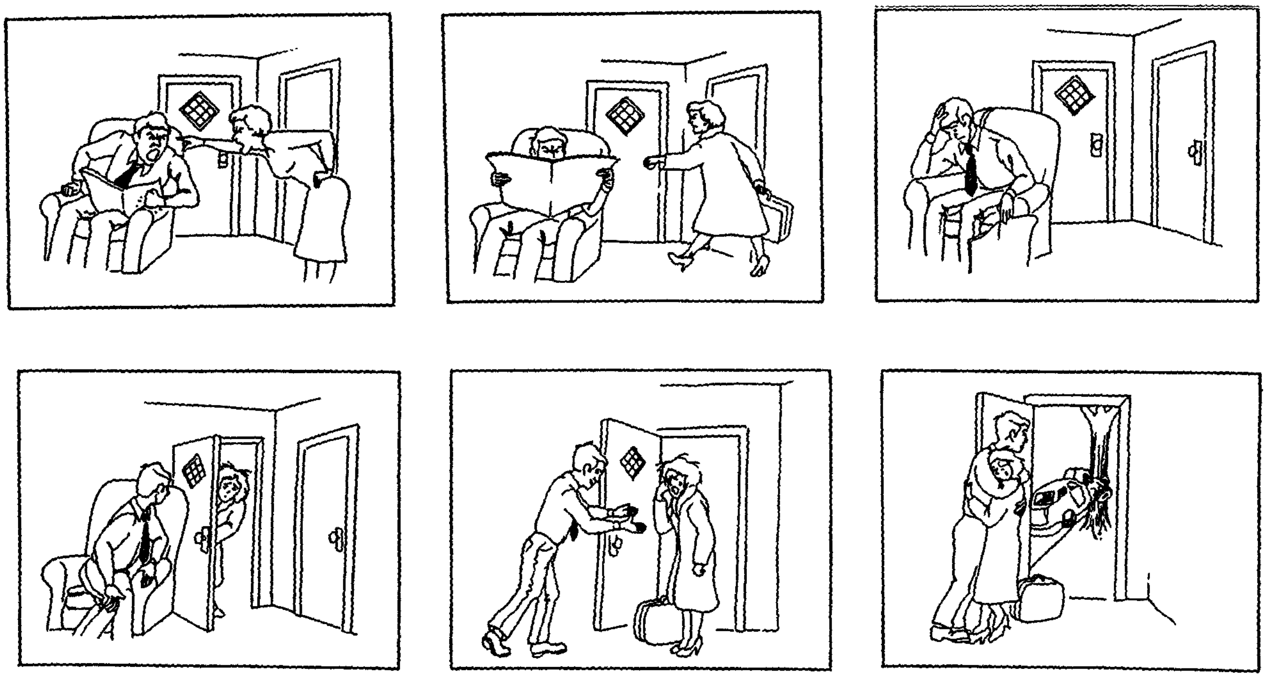
Appendix A

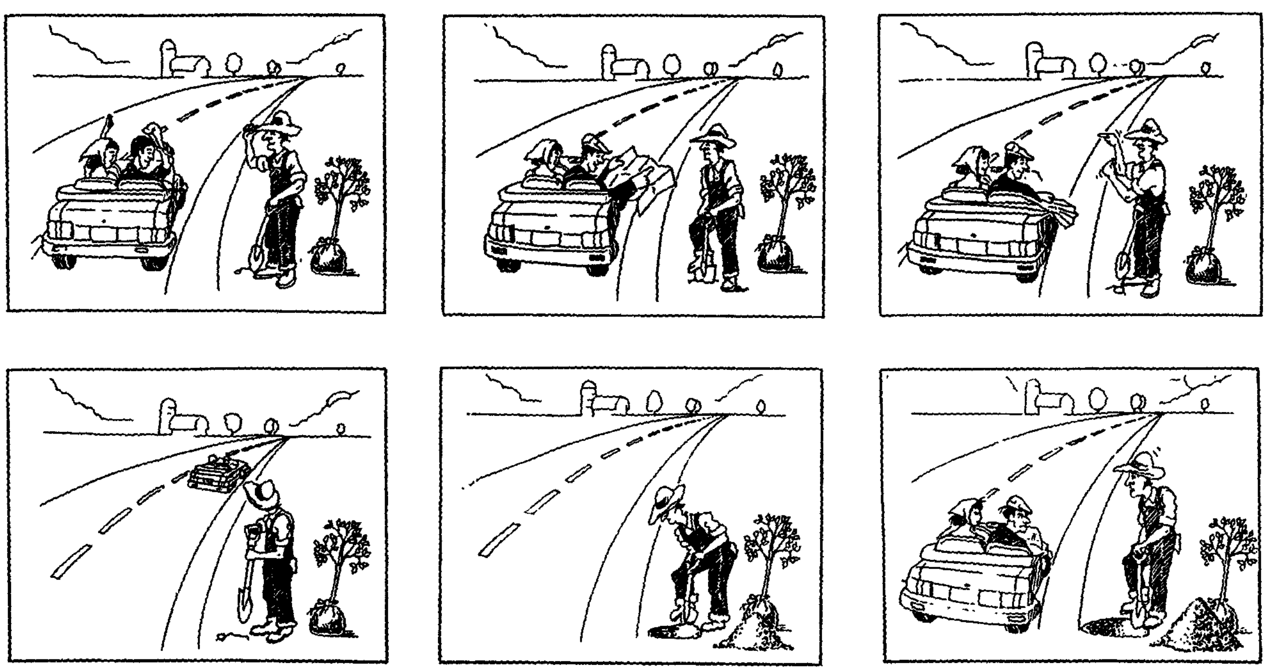
Two single pictures and two picture sequences drawn for this study. (Copyright 1992 by Robert H. Brookshire and Linda E. Nicholas)





Appendix A (continued)





Appendix B

Rules for scoring and counting words and correct information units (CIUs)

Prior to determining which words should be included in counts of words and correct information units, delete statements that are made before or after the speaker performs the task or suggest that the speaker is ready to begin or has finished the task and do not provide information about the picture(s) or topic itself. Such statements generally are not produced consistently by speakers from one session to another and are deleted to help stabilized counts across sessions.

- **I hope I can remember how I did this before.**
- **I'll start by saying this.**
- **I'm supposed to tell you about washing dishes.**
- **I'm ready to start.**
- **That's about it.**
- **I can't say any more.**
- **The end.**
- **That's about what our Sundays are like.**

These statements should be grammatically separate from discussion of the picture(s) or topic. The following first statements by a speaker would be included in the word count.

- In the first picture, the man is angry.
- Well first of all, there's a couple fighting.
- Okay, there's a man and a woman.
- Well now, here's a picture of a party.

This does not include commentary on the task or on the speaker's performance that occurs while the speaker is discussing the picture(s) or topic. (See 1.22 for rules about commentary.)

Instructions: Draw a horizontal line through the middle of words that are to be deleted prior to making decisions about the word count. (~~That's about it~~)

1.0. COUNTING WORDS

Definition: To be included in the word count, words must be intelligible in context to someone who knows the picture(s) or topic being discussed. Context refers to what the scorer knows about the picture(s) or topic and what the scorer knows from the speaker's prior words. Words do not have to be accurate, relevant, or informative relative to the picture(s) or topic being discussed to be included in the word count.

Instructions: Cross out with red Xs words that are not to be included in the word count.

RULES FOR COUNTING WORDS

1.1. DO NOT COUNT THE FOLLOWING

1.11. Words or partial words that are not intelligible in context to someone who knows the picture(s) or topic being discussed.

- He went to the **frampl**.
- That appears to be a **norble**.
- He had a **st . . . sn** . . . steak.

1.12. Nonword filler (**um, er, uh**). (See 1.23 and 1.24 for a rule dealing with filler words and phrases, interjections, and informal terms.)

1.2. COUNT THE FOLLOWING

1.21. All words that are intelligible in context. Count words that contain sound substitutions, omissions, distortions, or additions if the word is intelligible in context (**hiscup** for hiccup). If the incorrect production results is another real word that does not appear to be the target word, it is still included in the word count (**paper** for pepper).

1.22. Commentary on the task, on the speaker's performance, or on the speaker's experiences.

- **This is pretty hard.**
- **I can't think of that word.**
- **No, that's not right.**
- **My wife and I used to fight like that.**

1.23. Filler words and phrases (**you know, I mean, okay**). Do not count nonword filler. (See 1.12.)

1.24. Interjections (**oh, oh boy, wow, golly, gosh, gee, aha, hmm**) and informal terms (**uh-huh [affirmative], un-uh [negative], nope, yep, yeah**).

1.25. Common contractions or simplifications of words (gonna for **going to, sorta** for **sort of**, **em** for **them**). Contractions (both standard [**don't, he's**] and colloquial [**gonna, sorta**]) are counted as two words.

1.26. Each word in hyphenated words (**jack-in-the-box** = 4 words).

1.27. Each word in numbers (**twenty-two** = 2 words, **one hundred thirty-four** = 4 words, **nineteen fifty-five** = 3 words).

1.28. Compound words as one word (**pancake, cowboy**).

1.29. Each word in proper names (**Mary Smith, St. Paul, Mason City** = 2 words each).

1.30. Count acronyms as one word (**VA, VFW, TWA** = 1 word each).

2.0. COUNTING CORRECT INFORMATION UNITS (CIUs)

Definition: Correct information units are words that are **intelligible** in context, **accurate** in relation to the picture(s) or topic, and **relevant** to and **informative** about the content of the picture(s) or the topic. Words do not have to be used in a grammatically correct manner to be included in the correct information count. Each correct information unit consists of a single word and only words that have been included in the word count can be considered for inclusion in the correct information unit count.

Instructions: Put a diagonal pencilled slash through words that are not to be included in the correct information count ~~(man)~~.

RULES FOR COUNTING CIUs

2.1. DO NOT COUNT THE FOLLOWING

(In this section, words in **bold print** would not be counted as correct information units.)

2.11. Words that do not accurately portray what is in the picture(s) or that do not seem accurate in relation to the topic being discussed, such as incorrect names, pronouns, numbers, actions, etc. If a word reflects regional usage (such as calling the midday meal "dinner" in some areas), it is counted as a correct information unit. If grammatical incorrectness would lead to misunderstanding or uncertainty about the meaning of words, the grammatically incorrect words would not be counted as correct information units. (See 3.12 for examples of grammatically incorrect words that would be counted as correct information units.)

- The girl **is** riding her bike. (The picture shows a girl with a bike nearby which she may have been riding, but which she is not currently riding.)
- The **girl** is on a **ladder. She fell.** (The picture shows a boy on a stool who is tipping but has not fallen yet.)
- The **boys** and **girls** are arriving. (The picture shows only one boy and one girl arriving.)

If several people are involved in an action and only one of them is mentioned, the mentioned one is still counted as a correct information unit. This constitutes an incomplete description but not an inaccurate one.

The boy is arriving. (The picture shows a boy and a girl arriving.)

The man drove away. (The picture shows a couple driving away.)

2.12. Attempts to correct sound errors in words except for the final attempt.

- He put **paper popper** pepper on his food.
- She saw her with her **mass . . . mack**. . . mask.

2.13. Dead ends, false starts, or revisions in which the speaker begins an utterance but either revises it or leaves it uncompleted and uninformative with regard to the picture(s) or topic.

- **My si . . . no no not my sister . . . my fa . . .** with my wife.
- He goes over to her and **puts his** wants to give her a hug.
- He looks out and sees that **she had** the car ran into the tree.
- **The . . . the . . . that one oh forget it.**

- **In the hose In the mouse** in the house
- We go to **a party no I mean** a movie

If an utterance is incomplete, but some information about the picture(s) or topic has been given, count that information.

- The kitchen window was . . .

In this example, the words the kitchen window was would be counted as correct information units (if they meet the other criteria). Even though the entire statement was not completed, the words are informative.

Words that express some legitimate uncertainty or change in perception about characters, events, or settings in a picture are counted as correct information units (if they meet the other criteria). See 2.18 for further examples.

- Her dad or maybe a neighbor was in the tree.
- From the looks of the candles, he must be four. No there is another candle on the table so he must be five years old.

2.14. Repetition of words or ideas that do not add new information to the utterance, are not necessary for cohesion or grammatical correctness, and are not purposely used to intensify meaning.

- The **blue** truck was blue.
- The restaurant was a new one. **It was a new restaurant.**
- She was **cleaning** washing the dishes.

Such repetition of words or ideas can be separated by other counted words.

- The mother was very angry. The daughter was crying. **The mother was very mad.**

Exceptions:

(a) If the repeated words or ideas are necessary for cohesion, they are counted. • She went to the store. The store was closed.

(b) If words are repeated to achieve effect or to intensify a statement they are counted.

- The girl was very, very sad.
- They were fighting, really fighting.

(c) If repeated words are used to expand on previous information, they are counted.

- He put on a shoe . . . a left shoe.
- There were some people . . . a man and a woman.

2.15. The first use of a pronoun for which an unambiguous referent has not been provided. Subsequent uses of the pronoun for the same unspecified or ambiguous referent are counted as correct information units (if they meet the other criteria).

- **She** (no referent) was doing the dishes. I think she was daydreaming.

If an inaccurate referent is provided but it is clear that a pronoun refers back to it, the pronoun would be counted as a correct information unit.

- The fox (inaccurate referent) ate some of the cake and it was hiding.

2.16. Vague or nonspecific words or phrases that are not necessary for the grammatical completeness of a statement and for which the subject has not provided a clear referent and for which the subject could have provided a more specific word or phrase.

- The mother is drying one of those **things**.
- She gave him some **stuff**.
- He put **something** up to the tree but that **one** knocked it down.
- We had pancakes or scrambled eggs **or something like that.**
- I wash the glasses and plates **and so on.**

The words **"here"** and **"there"** frequently fall into this category.

- **Here** we have a boy.
- This **here** boy is crying.
- That mother **there** is doing dishes.
- There is a cat **here** and a dog **there**.
- The mother is **there**.
- She put them over **here**.
- She has a bike **there.**
- The cookies were up **there**.

The following are examples of uses of "here" and "there" that are necessary for the grammatical completeness of the statement and cannot be replaced by a more specific word. These uses of "here" and "there" would be counted as correct information units.

- There is a boy.
- Here comes the same couple.

The following is an example of a nonspecific word that is preceded by a clear referent and would be counted as a correct information unit.

- The boy opened the cupboard. The cookies were up there.

2.17. Conjunctive terms (particularly **so** and **then**) if they are used indiscriminately as filler or continuants rather than as cohesive ties to connect ideas.

- There is a man. **Then** there is a woman and **then** a cat.

When used cohesively, "then" indicates the temporal order or sequential organization of things or events.

- She had lunch and then she went to the store.
- When you go into my house you see the living room first, then the dining room, then the kitchen.

When used cohesively, "so" indicates a casual consequence.

- He was thirsty so he drank some juice.
- The mother was after the dog so the boy was crying.

2.18. Qualifiers and modifiers if they are used indiscriminately as filler or are used unnecessarily in descriptions of events, settings, or characters that are unambiguously pictured. The following examples concern unambiguously pictured information.

- **Apparently** this is a kitchen.
- **Evidently** the boy is on a stool.
- **I think that** the cat is in the tree.
- **It looks like** the man is up in the tree too.
- The boy is **sort of** crying and the dog is **kind of** hiding.
- **Of course**, the woman left in a huff.

When used informatively, qualifiers and modifiers suggest legitimate uncertainty on the part of the speaker about events, settings, or characters portrayed in the picture(s) or modify associated words in a meaningful way. The following examples concern ambiguously pictured information.

- Apparently this is a mother and her two children.
- I think she is his sister.
- It looks like he gave them the wrong directions.
- She must be daydreaming.
- He might be the girl's dad or maybe he's a neighbor.
- He is the father or a neighbor. I don't know which.
- He looks sort of sad.
- Evidently they went around in a circle.

2.19. Filler words and phrases (**you know, like, well, I mean, okay, oh well, anyway, yeah**), interjections when they do not convey information about the content of the picture(s) or topic (**oh, oh boy, wow, gosh, gee, golly, aha, hmm**), and tag questions (It is really smashed up, **Isn't it**).

2.20. The conjunction "**and**." "**And**" is never counted as a correct information unit because it is often used as filler and we have found that its use as filler cannot be discriminated reliably from its uses as a conjunction.

2.21. Commentary on the task and lead-in phrases that do not give information about the picture(s) or topic and are not necessary for the grammatical completeness of the statement.

- **These pictures are poorly drawn.**
- **This is kind of hard.**
- **In the first picture . . .**
- **As I said the last time,** she was upset.

2.22. Commentary on the subject's performance or personal experiences.

- **I can't think of the name of that.**
- **I can't say it.**
- **No, that's not right.**
- **My kids were always getting into trouble too.**
- **My wife and I used to fight like that.**
- They are fighting **but I don't know why.**

Some statements that contain personal information may be appropriate in procedural and personal information descriptions and, in such cases, they would be counted as correct information units (if they meet the other criteria).

See 3.16 for embellishments that are counted as correct information units.

See previous page for statements that are deleted before beginning the word and correct information unit counts.

3.1. COUNT THE FOLLOWING (if they meet all other criteria)
(In this section, words in bold print would be counted as correct information units.)

3.11. All words (nouns, adjectives, pronouns, verbs, adverbs, articles, prepositions, and conjunctions) that are intelligible in context, accurate in relation to the picture(s) or topic, and relevant to and informative about the content of the picture(s) or topic.

3.12. Words do not have to be used in a grammatically correct manner to be counted. Words that violate standard English grammar rules concerning appropriate verb tense and form, agreement in number between subject and predicate, agreement between articles and nouns, incorrect use of articles, and appropriate singular and plural forms are counted as correct information units unless these violations would lead to misunderstanding or uncertainty about the meaning of the words.
See 2.11 for examples of words that would not be counted as correct information units.

- The **firemans** are coming.
- The firemen **ain't** rescued them yet.
- Put some stamp on it.
- The friends **is** here.
- He **don't** look very happy.

3.13. Production of a word that results in another English word, if the production would be intelligible as the target word in context.

- He is standing on **a school** and it is tipping over.

3.14. The final attempt in a series of attempts to correct sound errors.

- He went to the musket . . . minuet . . . **market.**

3.15. Informal terms (**nope, yep, uh-huh, un-uh**) when they convey information about the content of the picture(s) or topic.

- She said "**Uh-huh**, I'll do it."

3.16. Words in embellishments that add to the events portrayed in the picture(s) or express a moral, if they are consistent with the situation or events portrayed. Words that express some legitimate uncertainty about characters, settings, or events in the pictures.

- **He's going to get hurt and his mom is going to be angry.**
- **Some days everything seems to go wrong.**
- **That looks like a nice way to spend a summer day.**
- **Sooner or later cats usually get stuck up a tree.**
- **Mothers sometimes get distracted and don't notice things.**
- **This is the one about the accident-prone family.**

However, see 2.22 for examples of extraneous commentary that may resemble embellishments, but are not counted.

3.17. Verbs and auxiliary verbs (**is, are, was, were, to, has, have, will, would, has been**, etc.) as two separate correct information units—one for the auxiliary verb and one for the main verb.

- His mom **is going to be angry**. (Each word in bold print is a correct information unit.)

3.18. Contractions [both standard (**won't**) and colloquial (**gonna**)] as two correct information units.

3.19. Each word in hyphenated words (**father-in-law, good-bye**).