CHAPTER

34

Comparison of Spontaneous and Elicited Oral-Expressive Language in Aphasia

Jan Avent Roberts
Robert T. Wertz

Conversational speech is disrupted in aphasia (Albert, Goodglass, Helm, Rubens, and Alexander, 1981), and it generally improves as aphasia improves (Prins, Snow, and Wagenaar, 1978). However, conversations are seldom evaluated systematically during assessment of aphasia.

Few comparisons of spontaneous and elicited aphasic language have been reported. Easterbrook, Brown, and Perera (1982) compared the speech of aphasic patients' conversation and their picture descriptions. Five patients were assessed at 6, 12, 18, and 24 weeks and at 9 and 19 months post-onset. The language samples were analyzed with the Language Assessment, Remediation, and Screening Procedure or LARSP (Crystal, 1982), and the investigators found that the elicited sample yielded more productive syntax than the conversational samples. However, a qualitative analysis revealed no apparent differences in the relative complexity and types of clauses and phrases used.

Albert and colleagues (1981) suggest the form, syntax, content, and semantics of conversational speech should be evaluated. Thus, we attempted to determine whether syntactic and semantic differences exist between conversational and elicited aphasic speech, and we attempted to determine whether differences might occur over time.

METHODS

Twenty patients who were aphasic subsequent to a single, left-hemisphere infarct were studied. Presence of aphasia was determined by the Porch Index of Communicative Ability (PICA) (Porch, 1967). All patients received 8 hours of treatment per week between 1 and 12 months post-onset. The patients' ages, education, and PICA overall percentile performances at 1 month post-onset are shown in Table 34-1. Patients' ages ranged from 41 to 78 years with a mean of 60, years of education ranged from 3 to 16 years with a mean of 10, and the PICA overall percentile ranged from 15 to 73 with a mean of 44.

Videotaped recordings of spontaneously produced oral-expressive language samples provided the data. Two language samples, conversation and PICA subtest I performance, were analyzed from each subject at 1, 3, 6, 9, and 12 months post-onset. The videotapes were randomized and coded to eliminate bias regarding the time post-onset. All videotaped conversations and PICA subtest I performances were then transcribed orthographically by conventional English spelling. Following transcription, the samples were analyzed by a T-unit analysis, syntactic well-formedness ratings, and sematic accuracy ratings.

A T-unit is defined as one main clause with all of the subordinate clauses or nonclausal structures attached to it (Hunt, 1970). For example,

TABLE 34-1. DESCRIPTIVE DATA AT 4 WEEKS POST-ONSET FOR AGE, EDUCATION, AND LANGUAGE SEVERITY FOR EACH PATIENT

Subject	*Age (years)*	*Education (years)*	*PICA overall percentile*
1	73	16	43
2	63	9	72
3	58	6	50
4	65	8	42
5	45	12	15
6	57	12.5	16
7	63	8	60
8	78	8	66
9	41	12	73
10	51	12	44
11	68	10	49
12	65	8	55
13	62	9	58
14	50	12	49
15	64	6	35
16	55	7	39
17	78	3	26
18	52	13	47
19	52	12	15
20	61	12	19

an utterance such as "I have things that upset me" would be considered as one T-unit with two clauses and six words. The T-unit measures obtained for the study were clauses per T-unit, words per T-unit, and words per clause.

Each T-unit was rated for percent of syntactic well-formedness. T-units were rated with a minus if the following were present: verb errors, obligatory word omissions, addition of extra lexical elements, incompleteness, plural errors, pronoun errors, word order errors, or awkward topicalization. Semantic accuracy percentages for each T-unit were similarly rated with a minus if the following were present: vague/empty vocabulary, given/new information errors, semantic or neologistic paraphasias, inaccurate information, ambiguous or without content, inappropriate lexical items, or incompleteness.

Five percent of the videotapes were selected randomly for determination of point-to-point individual word transcription interjudge and intrajudge reliability. Both interjudge and intrajudge reliability were above 90 percent. Ten percent of the data were scored to determine point-to-point word count, T-unit, and clause interjudge and intrajudge reliability. Interjudge and intrajudge reliability was above 85 percent.

The data were analyzed with two (language sample) by five (time) repeated measures analyses of variance. Language samples were conversation and PICA subtest I performance, and time was performance at 1, 3, 6, 9, and 12 months post-onset.

RESULTS

Mean syntactic and semantic performances varied over time on all measures in both conversation and PICA subtest I performance. Generally, performance on each measure increased between 1 and 12 months post-onset.

Syntactic complexity as measured by clauses per T-unit in conversation generally increased over time, while syntactic complexity in PICA subtest I initially increased and then declined after 6 months post-onset. Mean clauses per T-unit in conversation were 1.16 at 1 month, 1.35 at 3 months, 1.22 at 6 months, 1.29 at 9 months, and 1.36 at 12 months post-onset. Mean clauses per T-unit in PICA subtest I were 1.20, 1.28, 1.35, 1.24, and 1.21, respectively.

Clausal length generally increased over time in both conversation and PICA subtest I. Conversation performance in all time periods was higher than PICA performance. Mean word per clause in conversation over time was 4.5, 5.0, 5.1, 5.1, and 5.0 in PICA subtest I was 4.1, 4.3, 4.2, 4.4, and 4.7.

Utterance length generally increased over time in both conversation and PICA subtest I, and it was longer in conversation at all time periods than in PICA subtest I. Mean words per T-unit in conversation were 5.8, 6.9, 6.6, 6.9, and 6.9, and in PICA subtest I mean words per T-unit were 5.3, 5.6, 5.6, 5.7, and 5.9.

Syntactic well-formedness performance varied over time but showed increases in both samples from the initial 1 month post-onset performance. Mean percent of syntactic well-formedness in conversation was 64, 65, 73, 69, and 71. Mean percent of syntactic well-formedness in PICA subtest I was 64, 67, 66 76, and 72.

Semantic accuracy also increased over time in both samples but was generally better in conversation than in PICA subtest I. Mean semantic accuracy performance over time in conversation was 68, 76, 82, 82, and 84 and in PICA subtest I was 56, 80, 73, 77, and 80.

The two (sample) by five (time) repeated measures analyses of variance were performed on each of the syntactic and semantic language performances measures to test for differences between conversation and PICA subtest I. Each measure was examined for a main effect of sample source and effect of sample within each of the five time periods.

Statistically significant main effects were found between sample sources on words per clause ($F[1,19] = 4.61, p < .05$), words per T-unit ($F[1,19] = 5.43, p < .03$), syntactic well-formedness ($F[1,16] = 7.26, p < .02$), and semantic accuracy ($F[1,16] = 11.80, p = .01$). These results indicate that utterance length and clausal length were statistically longer in conversation than in PICA subtest I. Mean utterance length was 6.63 in conversation and 5.62 in PICA subtest I. Mean clausal length was 4.92 in conversation and 4.32 in PICA subtest I. Semantic accuracy was statistically significantly better in conversation, 78.55, than in PICA subtest I, 73.83. Conversely, syntactic well-formedness was statistically significantly better in PICA subtest I performance. Mean pica subtest I performance was 69.04 percent accurate. Mean conversation performance was 68.60.

Statistically significant differences were also found for sample within time for words per clause ($F[5,95] = 5.46, p = .01$), words per T-unit ($F[5,95] = 6.94, p = .01$), and semantic accuracy ($F[5,80] = 7.45, p = .01$). These results indicate that statistically significant differences exist between conversation and PICA subtest I at specific time periods.

DISCUSSION

Our results indicate that differences exist between conversational and elicited language samples. Overall utterance and clausal lengths were longer in conversation. Semantic accuracy was better in conversation, but syntactic accuracy was better in PICA subtest I.

Differences between conversation and PICA subtest I performance on utterance and clausal lengths and semantic accuracy also occurred at specific points in time post-onset. Because performance varied among measures at different points in time post-onset, it may be necessary to evaluate both spontaneous and elicited performance periodically, at least, during the first year post-onset.

Differences in syntactic complexity between the two samples were evident, but no statistically significant differences were detected. Similar results were obtained by Glosser, Wiener, and Kaplan (1988). These investigators compared verbal complexity in narration and picture descriptions. They found that the picture task tended to elicit lower verbal complexity performance than the narration task, but the differences were not statistically significant. The lack of statistically significant differences in syntactic complexity may indicate that most language samples are an adequate predictor of overall syntactic complexity performance. However, it may also indicate that currently used measures of syntactic complexity are insensitive to differences between or among samples.

Standardized, elicited testing of aphasic patients is important for diagnosis, focusing treatment, and measuring change. However, we observed that elicited verbal performance may either under- or overestimate conversational syntactic and semantic language skills at different points in time. Because aphasic people converse more than they take standardized language tests, we may want to sample and analyze conversational skills as rigorously and precisely as we employ standardized tests.

REFERENCES

Albert, M. L., Goodglass, H., Helm, N. A., Rubens, A. B., and Alexander, M. P. (1981). *Clinical aspects of dysphasia.* New York: Springer.

Crystal, D. (1982). *Profiling linguistic disability.* London: Edward Arnold.

Easterbrook, A., Brown, B. B., and Perera, K. (1982). A comparison of the speech of adult aphasic subjects in spontaneous and structured interactions. *British Journal of Disorders of Communication, 17,* 93–107.

Glosser, G., Wiener, M. and Kaplan, E. (1978). Variation in aphasic language behavior. *Journal of Speech and Hearing Disorders, 53,* 115–124.

Hunt, K. (1970). *Syntactic maturity in school children and adults.* Monograph of the Society of Research in Child Development.

Porch, B. E. (1967). *The Porch Index of Communicative Ability.* Palo Alto: Consulting Psychologists Press.

Prins, R. S., Snow, C. E., and Wagenaar, E. (1978). Recovery from aphasia: Spontaneous speech versus language comprehension. *Brain and Language, 6,* 192–211.

DISCUSSION

Q = question; A = answer; C = comments.

Q. One of the things we've noticed with older children and have not looked at systematically with aphasic adults is not so much the complexity in syntactic formations but the richness of their syntactic repertoire. Now one of the questions that occurred to me that you might be overlooking and might have gotten a gut-reaction for is that the patients as they recover may increase their syntactic complexity but reduce their syntactic repertoire. In other words, they become comfortable with certain frames. Now Natalie Waterson has written about this in development processes, and I'm just wondering if you would comment on that?

A. What do you mean by "repertoire"?

Q. What I mean is the number of different types of syntactic utterances that they make. We call that a fragile expressive language system. Within their syntactic performance they're very good but they don't use a lot of different types of syntactic constructions. So, getting better by narrowing the scope of performance. So I'm wondering if that might be some part of what you might be seeing?

A. A couple of things come to mind. Dan Kempler looked at variety of syntax in Alzheimer patients, and that's one way to go about looking at the variety, and you may find there are some constructions that patients are more comfortable with and use more often. Instead of looking at amount, it may be more important to look at type of syntax and capitalize on what patients use more comfortably.

Q. Were the patients treated, and if so what was the treatment? What I'm wondering is, since you're looking at this information over time, was the treatment directed at these types of deficits? And could you speculate on the impact of that on a time-protracted analysis such as yours?

A. The patients received 8 hours of treatment per week, and some received direct, individual treatment and some received group treatment. There were differences in treatment, but we didn't treat complexity, so no one received direct treatment for that. There is a study by Shewan that just came out and reported on complexity. Complexity is an interesting variable because it tends to vary in normal and non-brain-damaged individuals. So, I don't think we understand all of the components of complexity. It's something that we can treat, I think, but we don't typically treat it.

C. I would like to make a comment about syntactic complexity. I think it is a very difficult issue, and I feel that syntactic complexity per se doesn't mean much. I think it is closely related to the content and the type of task you give to the patient, and if you're looking at changes in syntactic complexity over time, I think you can see it if you choose your tasks wisely, and in some of the tasks — for example in some types of expository discourse — you can see these increases in syntactic complexity, while in others it might not be noticeable at all. So I feel that syntax is just a tool and is closely related to semantics.

A. I agree. I think you will see differences depending on the task. My concern about complexity is that there is a tendency in the literature to assume that you can't treat syntax. Performance in the acute phase of stroke will be tha same in the chronic phase. But, if you see changes in syntax regardless of whether it is from the task you employ or by treating it, then I think it is a viable treatment goal and we shouldn't assume that syntax can't improve over time.

Q. We've seen changes in syntactic complexity across types of discourse going from procedural, to narrative, to summaries of narratives. Sometimes aphasic patients will produce more complex syntax on a summary, for example, than is actually called for or normals would produce. And, in that case, having more complex syntax is really more abnormal. So, as you say, it's good that they can produce it, and maybe that could be used to facilitate more complex syntax in appropriate contexts. My question is, without any normative data where you are comparing aphasic syntactic complexity to normal syntactic complexity across these as yet poorly studied varieties of discourse, it's difficult to interpret these types of changes in a meaningful way. Did you have any normal data?

A. Not yet. That's a very good point. There is some literature from the 1960s and early 1970s that documented syntactic changes in children up to a certain age level. But you are right that there is no "goodness of fit" for an analysis such as this. And, certainly we can see from our data that some fluent aphasic patients produced average clauses per T-unit that were quite high, and I think what we would find is that these averages would be higher than non-brain-damaged adults would produce. So, definitely, we need that normative data.

Q. Yes, I think we vary many aspects of discourse, including syntactic complexity, going from one type of discourse to another and probably in different contexts. So there's a tremendous amount of work to be done to figure out what non-brain-damaged people do. And, this goes back to what we were talking about last night regarding how are we to interpret all of these subtle changes that vary from context to context?

A. And, in fact, it might be variation that we'd like to see — that there is less complexity in one task than in another because that's probably what non-brain-damaged people do. And you are right, there's a lot of work to be done.

Q. What was your conversational task, and do you think you got a lot of random variability because it's a more loosely organized task? And, could you get a stable baseline with a conversational speech sample stimulus?

A. I don't know about the stability of it. That's a really good point. The conversational topics were biographical in nature, for example, when

A. Not in a preliminary study such as this, but I think it would be a good follow-up study to group patients with similar severity levels possibly to reduce variability.

Q. Do you think severity it analogous to fluency or something like that? Is that what you mean?

A. You could group patients according to severity or according to aphasia subtypes.