Investigation of the Sentence Hierarchy of the Helm Elicited Language Program for Syntax Stimulation

Anthony P. Salvatore and Mary Jane Trunzo
Arizona Speech-Language and Learning Clinic, Phoenix, Arizona

Pat Holtzapple and Lisa Graham
Veterans Administration Medical Center, Phoenix, Arizona

The need to develop effective treatment procedures for adults who exhibit acquired verbal production disorders is clear. Helm-Estabrooks (1981), recently published the Helm Elicited Language Program for Syntax Stimulation, commonly referred to as the HELPSS, and recommends it as an effective treatment program for improving the use of syntactic form in agrammatic and paragrammatic individuals. The HELPSS treatment package is predicated on the hierarchical arrangement of the sentence types reported by Gleason, Goodglass, Green, Ackerman & Hyde (1975). In critiquing the Gleason, et. al., study, Noll (1981) notes that "there was considerable variation among the subjects in the relative difficulty of the sentence types" (p. 489). Furthermore, our experience with the program led us to conclude that the HELPSS is not explicit in defining the nature of an accurate response.

PURPOSE

The data presented in this paper will deal with only two of the clinically pertinent questions that arose from our experience with the program as part of a long term research effort. (1) How do we generate patient-specific hierarchies? (2) What is an accurate response on the HELPSS?

METHOD

Subjects. Seven aphasic adults participated in our study. Table 1 presents biographical data. All patients were judged to be aphasic by a neurologist and a certified speech and language pathologist. The subjects were recruited from three treatment centers in the Phoenix metropolitan area. Table 2 shows the pretreatment test results on the Porch Index of Communicative Ability, the PICA (Porch, 1967), and tests of auditory comprehension.

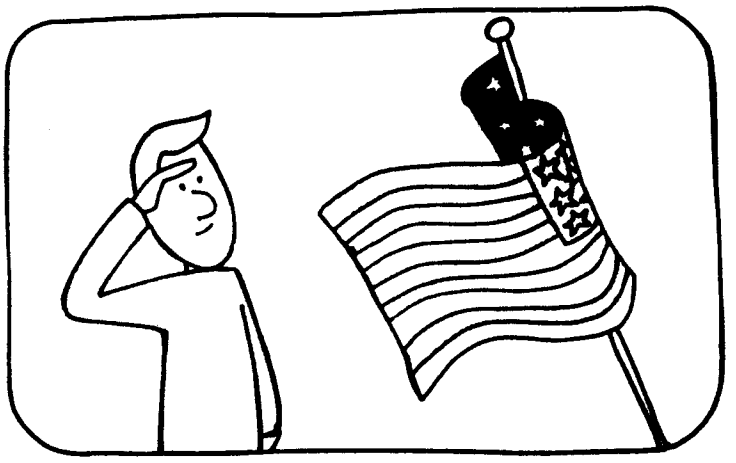
Table 1. Subject description.

SUBJECTS	AGE	EDUCATION	ETIOLOGY	TYPE OF APHASIA	TIME POST ONSET
1	54	HS	CVA	Anomic	3 weeks
2	69	HS	CVA	Broca	52 months
3	61	2 years college	CVA	Conduction	21 months
4	66	HS	CVA	Broca	8 months
5	34	College graduate	CVA	Wernicke	1 month
6	41	College graduate	CVA	Broca	39 months
7	27	HS	Closed Head injury	Broca	68 months

Table 2. Pretreatment test results.

SUBJECTS	PICA PERCENTILES OVERALL	I	VI	X	XII	AUDITORY COMPREHENSION TESTS TOKEN TEST	NCCEA(TEST 1
1	58	63	63	38	66	20% 6/30 correct	--
2	53	50	35	50	50	--	--
3	62	40	70/99	65/99	25	83% 52/62 correct	--
4	23	24/29	2	2	33	Failed screening	--
5	39	51	29	48/54	70/74	--	74% 29/39 co
6	56	52	33	37	57/58	--	52% 10/19 co
7	60	51	38/40	35	60	--	38% 15/39 co

Materials. Instructions for the picture and story presentation were administered in the manner described in the HELPSS manual. Figure 1 shows an example of the HELPSS stimulus materials.



PROCEDURE

The administration manual offers the following suggestion: "training is best based on what is known about the performance of the population to be treated" (p.2). We interpret this statement to mean that the clinician would start every patient at Sentence Type (ST) 1 and work sequentially through the hierarchy Helm-Estabrooks (1981) does acknowledge in the manual that "no single approach can or should be strictly applied to every client" (p.9). She further states that "It is axiomatic in aphasia therapy that the most profitable approach is to begin where the client has the greatest chance of making a successful response ..." (p.2.). Yet, given a patient whose success does not conform to the a

priori 1 through 11 hierarchy, Helm-Estabrooks offers no suggestions as to how the hierarchy might be modified.

Thus evolved our first question, "How do we generate patient-specific herarchies?" We decided to let each patient's performance dictate the nature of their hierarchy. We hoped that each patient's performance across three baseline sessions would provide us with a clear indication of which STs were easiest for him or her--the STs which yielded the greatest number of correct responses. Furthermore, the baseline data would later provide a referent for judging the efficacy of the treatment package.

Each patient was administered a sample of the sentence types at each level of complexity, A and B, over three consecutive sessions ranging from three days to three weeks to establish a baseline performance. Sentence Types were administered in the hierarchical order of HELPSS (Table 3).

Table 3. HELPSS hierarchy of sentence types.

1.	Imperative Intransitive
2.	Imperative Transitive
3.	WH - Interrogative
4.	Declarative Transitive
5.	Declarative Intransitive
6.	Comparative
7.	Passive
8.	Yes - No Questions
9.	Direct & Indirect Object
10.	Embedded Sentences
11.	Future

The first five picture-sentence items from each ST (except 1A) were designated as baseline stimuli for Level A. The last five picture-sentence items were used as baseline stimuli for Level B. There were 50 level A stimulus items and 55 Level B stimulus items. A total of 315 verbal responses were produced by each patient, and verbatim transcriptions of their 165 Level B responses were made for purposes of analysis.

As we administered the baseline stimuli, it became apparent that there was poor interexaminer scoring reliability. Thus evolved our second question: "What is an accurate response on the HELPSS?" Reference to the manual suggests that the purpose of the program is to increase the frequency of use of the sentence structures trained. Therefore, a reasonable interpretation of the manual is that the STRUCTURE of the utterance is the dependent variable to be scored. The scoring system described in the HELPSS manual is:

1.0 Fully correct response
0.5 Self-corrected response
0.0 Incomplete or incorrect response

Helm-Estabrooks (1981) described a correct response at each level, A and B, as follows:

"Level A requires the client to repeat the stimulus sentence that the clinician presents within the story...(p. 9).

"Level B requires the client to complete the story presented by the clinician with the appropriate sentence." (p. 10).

The only example of an incorrect response offered in the manual (p. 7) is:

Target Response	Subject's Response
He is taller.	He is tallest.

As you can imagine, these definitions of accuracy preclude agreement among clinicians. To improve our own interexaminer scoring reliability we developed the following post hoc experimental categorization procedure, affectionately called the Sonoran System (Appendix A). The Sonoran system is based on a match or mismatch of the patient's response to the target ST (Table 4). This binary system was concieved to assess syntactic and semantic accuracy in the patient's responses. Based on the Sonoran System for categorizing responses, an accurate response could be any one or combination of the five categories, S_1, O_1, N_1, O_2, R (Appendix A), depending on your interest in syntax, semantics or a combination of both.

Table 4. Binary description of syntactic and/or semantic match.

	Syntactic Match	Semantic Match
Potential Accuracy Categories		
S_1	+	+
O_1	+	+
N_1	+	+
O_2	-	+
R	+	-

RESULTS

Our first question was "How do we generate patient-specific hierarchies?" We had hoped that each patient would tell us that by his baseline performance. What we found was that each patient's success on the different sentence types varied. Performance of each patient varied not only from baseline to baseline, but the hierarchies generated by each patient differed according to which categories were included to define an accurate response as shown in Table 5.

We pooled the data from the seven patients (Table 6) and developed a hierarchy of performance for the group. We discovered that the group performan did not match the hierarchy presented in the HELPSS manual. The sentence hierarchy shifted, based on the definition of accuracy. Thus, our group data do not support the hierarchy described by HELPSS. This is strong evidence that each patient's hierarchy must be established individually to determine where to begin treatment. In addition to establishing individual hierarchies, the same data provide us with a referent against which to measure the efficacy of the treatment program.

Response stability was not observed across baseline sessions or across patients. Five of the seven patients showed positively accelerating performanc across the three sessions. These five patients had suffered a cerebrovascular

Table 5. Hierarchies Determined By 4 Different Definitions Of Accuracy Based On The Sonoran System.

HELPSS Sentence Type	S_1	$S_1O_1N_1$	$S_1O_1N_1O_2$	$S_1O_1N_1$ - R
Subject 1				
1	11	11	5*	11
2	3	8	11	8
3	7	3	7	4
4	8	4	3	6
5	5	5	4	9
6	6	7	6	3
7	9	9	8	5
8	zero	6	1	7
9	NR	1	9	2
10	↓	2	10	1
11		10	2	10
Subject 2				
1	NR	1	5	1
2	↓	NR	2	4
3		↓	1	NR
4			NR	↓
5			↓	
6				
7				
8				
9				
10				
11				
Subject 3				
1	1	1	1	1
2	NR	2	2	2
3	↓	NR	7	4
4		↓	10	8
5			4	11
6			6	NR
7			8	↓
8			NR	
9			↓	
10				
11				

* | = Tie

accident (CVA). Of the two patients who showed stable performance, one was an individual with a closed head injury and the other had a CVA.

Table 6. Sentence type hierarchy for each definition of accuracy across three baseline sessions for seven patients.

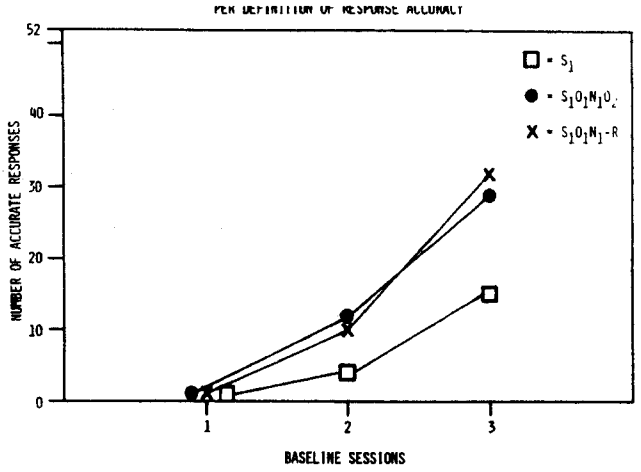
	HELPSS	S_1	$S_1O_1N_1$	$S_1O_1N_1O_2$	$S_1O_1N_1$ - R
EASY	1	11	5	5	8
	2	3	8	1	5
	3	8	11	6	1
	4	6	1	8	6
	5	1	6	11	11
	6	5	3	2	2
	7	2	4	4	3
	8	7	9	9	4
	9	4	2	3	9
	10	9	7	10	7
DIFFICULT	11	10	10	7	10

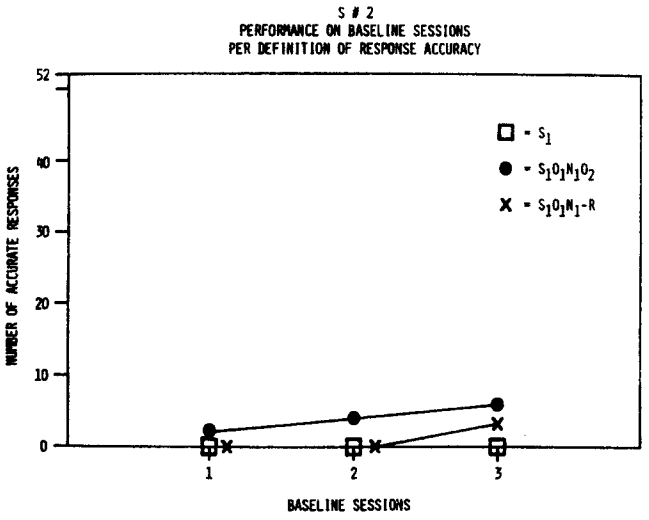
The patients improved over time regardless of the definition of accuracy. Individual patient data are shown in Figure 2. The CVA patients consistently demonstrated improvement across sessions. For this reason, it is imperative that data be collected by means of baseline measures. Without stable baseline performance, claims of program efficacy cannot be justified.

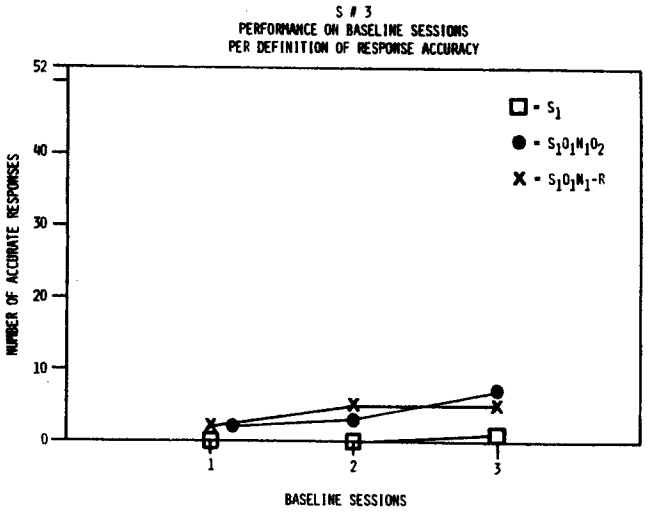
DISCUSSION

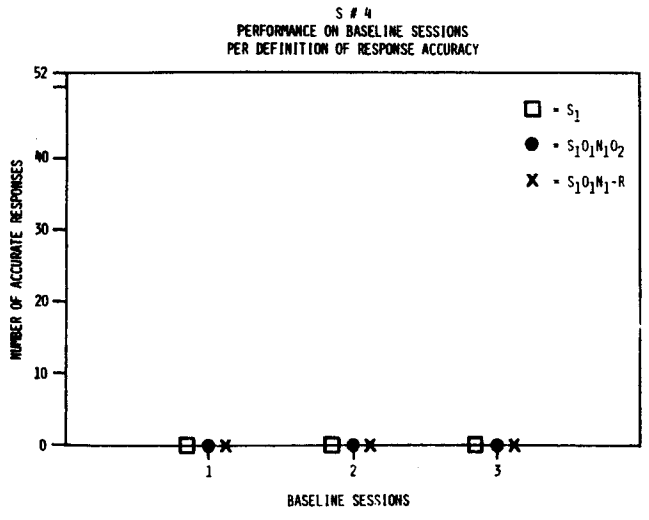
In answer to our first question "How do we generate patient-specific hierarchies?", it is clear that each patient's hierarchy must be defined by his own baseline performance. In response to the second question, "What is an accurate response on the HELPSS?", we offer the following. When developing an operational definition for response accuracy, the clinician looking for improved use of syntax will have different criteria for a correct response than the clinician whose purpose is increased functional communication skills. Whatever intentions the clinician has in using the HELPSS, an operational definition of response accuracy must be stated to insure intra- and inter-examiner reliability.

Finally, improved performance over repeated sessions dictates the need for baseline assessment. A baseline measurement is not one data point. As Sidman (1960) has stated, "A behavioral baseline is not some idealized state of behavior inferred from the performance of a group of individuals by means of a statistical averaging process. It is the continuous, and continuing, performance of









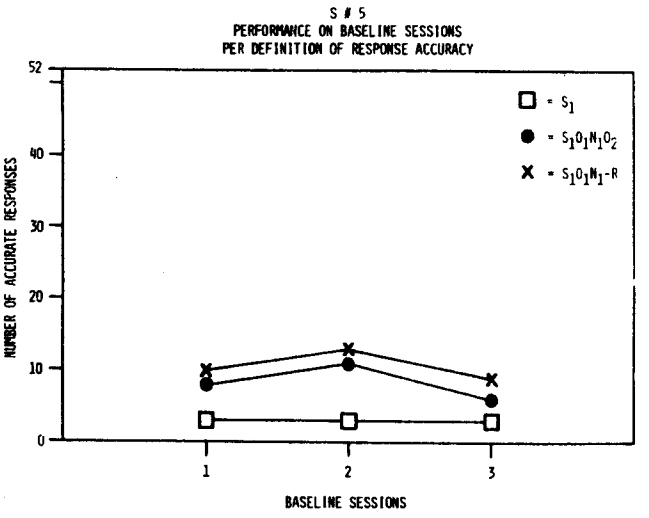


Figure 2. Subjects' performance in Baseline Sessions per definition of response accuracy. (Figure continued on next page.)

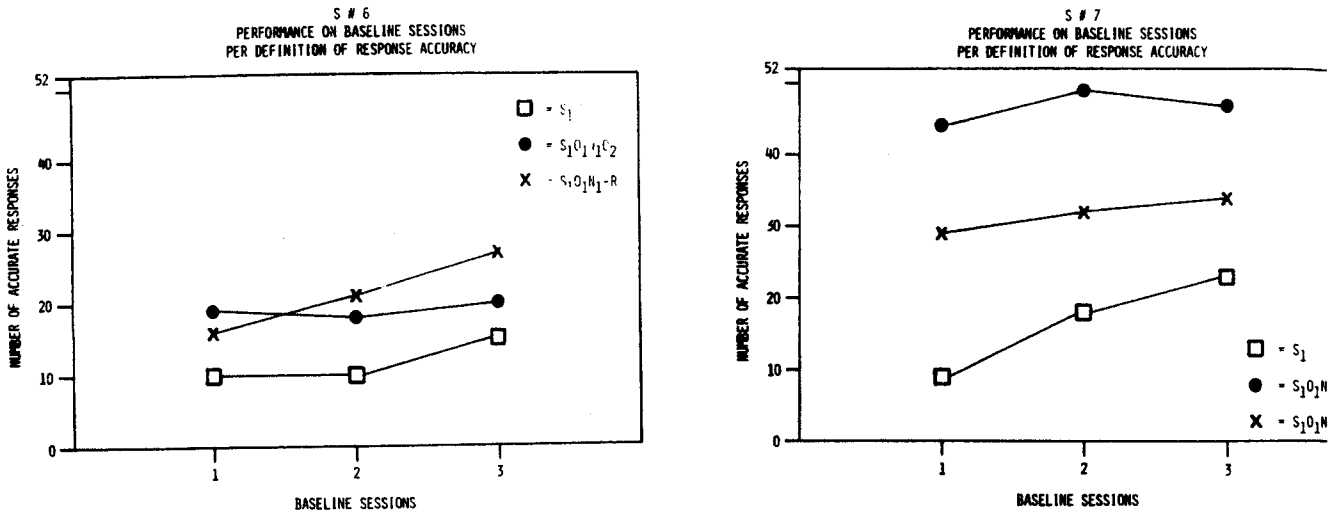


Figure 2. (cont) Subjects' performance in Baseline Sessions per definition of response accuracy.

a single individual" (p. 409). We sincerely hope that our efforts, those of the author of the program, and those of other clinical researchers, will lead to definable and measurable procedures for using the HELPSS as a relevant clinical tool.

APPENDIX A

THE SONORAN SYSTEM: AN EXPERIMENTAL CATEGORIZATION PROCEDURE

S_1 Replication of HELPSS target.

O_1 HELPSS Syntax match, Semantic match (lexical substitution).

	Target (T)	Response (R)
1B10	T: Yell	
Relativization (reverse order)		R: Holler
9B11	T: He throws the dog a bone.	
	R: He throws a bone to the dog.	

N_1 HELPSS syntax match, Semantic match

1. Additional information
   1B14 T: Run R: Run faster

2. Carrier phrase
   4B4 T: He eats his lunch.
   R: I tell them that he eats lunch.

3. Additional information and lexical substitution
   8B2 T: Did you buy the paper?
   R: He says, did you get me a copy of the paper?

O_2 HELPSS Syntax mismatch, syntax complete, Semantic match
   1B15 T: Salute R: Salute the flag

R HELPSS Syntax match, Semantic msimatch (related error, wrong information)
   SB4 T: Open the window R: Open the door

A HELPSS Syntax mismatch, Semantic mismatch (related error). Appropriate attempt to answer the question
   11B1 T: He will walk R: He walks

N_2 HELPSS Syntax mismatch, Semantic mismatch (unrelated error).
   10B3 T: She wanted him to be popular
   R: lost and found

S_2 Syntactic, Semantic self-correction to a $S_1O_1N_1O_2R$. Does not include articulation self-correction.

Y Rejection

S_3 Unintelligible

T No response

E Echo

M HELPSS Syntax mismatch, Semantic mismatch (related error). Response is comment on auditory or visual stimuli; not an appropriate attempt to answer the question.

1B11 T: Watch out R: Boards

REFERENCES

DeRenzi, E., and Vignolo, L.A., The Token Test: A sensitive test to detect receptive disturbances in aphasia, Brain, 85, 665-678, 1962.

Gleason, B.J., Goodglass, H., Green, E., Ackerman, N., and Hyde, M.R., The Retrieval of Syntax in Broca's Aphasia, Brain and Language, 2, 451-471, 1975.

Helm-Estabrooks, N., Helm Elicited Language Program for Syntax Stimulation. Austin, Texas: Exceptional Resources Inc., 1981.

Helm-Estabrooks, N., Fitzpatrick, P., Barresi, B., Response of an agrammatic patient to a syntax stimulation program for aphasia. Journal of Speech and Hearing Disorders, 46, 422-427, 1981.

Noll, J.D., ASHA, July, 489-490, 1982.

Porch, B.E., Porch Index of Communicative Ability, Volume 1, Theory and Development. Palo Alto, California: Consulting Psychologist Press, 1967.

Sidman, M., Tactics of Scientific Research: Evaluating Experimental Data in Psychology. New York: Basic Books, 1960.

Spreen, O., and Benton, A.L., Neurosensory Center Comprehensive Examination for Aphasia (NCCEA), 1977 Revision. Victoria, British Columbia; Neuropsycholo Laboratory, University of Victoria, 1969.

DISCUSSION

1. Q: What is the intrasubject reliability on the hierarchies?
   A: We found no significant correlation between the HELPSS herarchy and the group hierarchy, nor did we find significant correlations across the hierarchies for each baseline for each subject.

2. Q: What treatment decisions does one make if the hierarchies change across baseline sessions?
   A: That is the subject of a forthcoming paper. If we have time, I would like to come back to that question.

3. Q: Even though the hierarchies were not consistent across sessions, did you find some sentence types that tended to be easier than others?
   A: Yes, the eleven sentence types fell into two or three and sometimes four groups of sentences based on a subject's performance. Clinically speaking, we must consider how important separating out sentence types based on differences of 1, 2, or 3 correct responses between sentence types is. That is, were the differences we observed between the sentence types clinically significant?

4. Q: In your judgement of an accurate or allowable response, did you have any apractic-aphasic errors in your patients so that you would have to start them off with some initial cuing?
   A: If a response was totally unitelligible, we scored it unintelligible. We did not give any phonemic cues. We followed the procedure for administration described in the manual.

5. Q: Based on your experience with this procedure, do you have a better idea what type of patient would benefit from it?

A: Our feeling is that auditory comprehension skills of the patient are important. There were two patients (#4, #7) who had very poor auditory comprehension on the Token Test and did poorly across the baseline sessions. Another patient (#1) performed poorly on the Token Test and did rather well on the baseline material. The role of auditory comprehension and success on the HELPSS procedure must be defined.

6. Q: Were all these patients agrammatic?
   A: No. we had a mix of etiologies.

7. Q: Isn't this program just designed for agrammatic patients?
   A: No, not according to the administration manual. The author suggests that the procedure may also be helpful to paragrammatic patients. We used a mix of patients to facilitate the identification of patients who might respond appropriately to this program. The study reported by Helm-Estabrooks in the Journal of Speech and Hearing Disorders, 1982, reported on an agrammatic patient. However, patient #5 is paragrammatic and showed improvement across baseline sessions and, although not reported here, improved during treatment.

8. Q: Is it your feeling that hierarchies, at least in this realm, might never be relevant? Rather, that something like this should be presented as just another task and you determine the hierarchy for your individual patients on a particular day, or will we come to some kind of hierarchy but a different kind, be it as groupings or with different patients?
   A: It would be nice if we could find reasonable hierarchies. It would make our jobs a lot easier. If we don't we're going to spend time trying to determine what is appropriate for each patient.

9. Q; Did it take a lot of time?
   A: Yes, collecting the data was not time consuming but developing the SONORAN System to analyze the data took a lot of time.
   On a positive note, Helm-Estabrooks has taken the time to collect these stimulus materials, arrange them in an orderly fashion, and provide, for the most part, some nice pictures. However, there are some sentence-picture combinations that allow a tremendous number of possible responses from the patient. We feel the original Gleason et. al., stimuli are more appropriate. The sentence types used in the HELPSS differ somewhat from those used in the Gleason et. al., study. Thus it is questionable if one can extrapolate a hierarchy from the Gleason et. al., study to the HELPSS. For example, the Declarative Intransitive:

   Gleason et. al.,: Dogs always chase cats. A dog is in the street. A cat comes along. What happens?
   Target: "The dog chases a cat" or "dog chase cat".
   HELPSS: People ask what my brother does for the construction company. I tell them what?
   Target: He builds houses.

10. Q: To what extent was some of the variability across subjects due to the fact that you included in your study subjects for whom the program was not designed?

A: What I suggested in my earlier comment was that paragrammatic patients are not excluded from the program according to the manual. Two patients are reviewed in the manual, one agrammatic and one paragrammatic, suggesting that the latter may benefit from the procedure.

11. C: There is a real danger in the whole program as proposed because it is based on a very small data base. Furthermore, studies like the Gleason study wash out the problems of individual differences.

12. Q: Did the Wernicke's patient who had poor auditory comprehension do well on the baseline measures?
A: Yes. Our expectations that good auditory comprehension was necessary was not fully satisfied. A couple of patients with good auditory comprehension did not do as well as a patient with poor auditory comprehension.

13. Q: Do you have a greater problem with the manual or the program?
A: The program is nothing more than what is observed in the manual. I have difficulty defining the program. I am sure Nancy Helm-Estabrooks can and would, but picking up the manual and trying to use it is very difficult.

14. Q: Do you have any feeling about the method of eliciting the response? That is, what about the story presentation method?
A: Can I make a generalization? When we got to Level B and the target response wasn't limited by the antecedent story, the patient would either pick out some portion of the story to repeat or describe the picture. Is that a fair generalization. Yes!