Experimental Analysis of a Syntax Stimulation Training Procedure

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

Two years ago we presented a report (Salvatore, Holtzapple, Trunzo and Graham, 1983) to this conference which raised what we thought were important clinical issues pertaining to the clinical application of the Helms Elicited Language Program for Syntax Stimulation (1981) (HELPSS). In that report we stated that the establishment of a hierarchy of difficulty should be based on an individual's performance and not on an a priori hierarchy such as that proposed in the HELPSS manual. We substantiated that suggestion when we reported that none of seven subjects studied conformed to the a priori hierarchy presented in the manual.

Next we suggested that the scoring criteria offered in the manual were not sensitive to the verbal responses emitted by the patients during baseline testing. We offered an alternative scoring system that more accurately tracked the verbal responses of the subject. Finally we demonstrated that performance varied across three consecutive sessions during measurement of baseline performance, thereby questioning case study reports dealing with the efficacy of the HELPSS procedure.

Today's report deals with the use of the HELPSS stimuli in conjunction with a training procedure designed to treat grammatic speech deficits in brain damaged adults. The following questions are reported on.

1. Is the HELPSS training procedure effective in producing accurate grammatical verbal responses to the training stimuli?
2. Is the HELPSS training procedure effective in producing accurate verbal responses to untrained stimuli?

METHOD

Subjects. Three aphasic adults participated in this study. Table 1 presents subjects' biographic data. All subjects were labeled aphasic by a neurologist and certified speech and language pathologist. The subjects were recruited from three treatment centers in the Phoenix metropolitan area.

Table 1. Subject characteristics.

Subject #	Age	Education	Etiology	Type of Speech	Time Post Onset	PICA OA
1	31	16 years	Open Head Injury	Nonfluent	9 months	75 %ile
2	62	12 years	Multi-CVA's	Nonfluent	3 months	81 %ile
3	51	6 years	Closed Head Injury	Fluent	14 months	65 %ile

Materials. Sentence and story materials, along with the sentence hierarchy and picture stimuli were taken from the HELPSS manual (Figure 1).

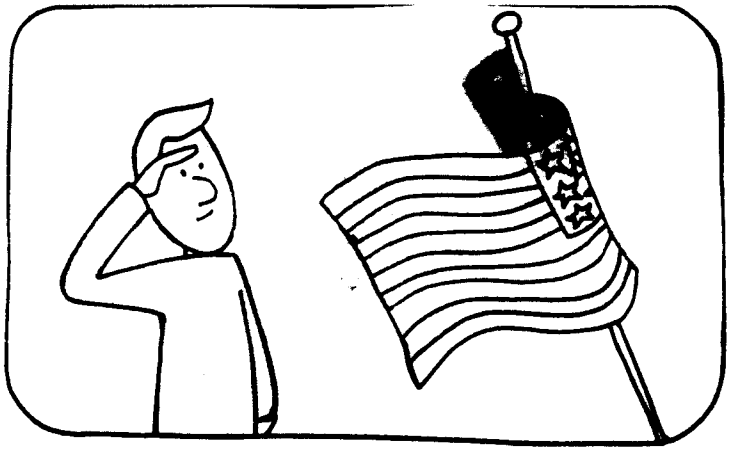

Figure 1. Example of a stimulus item from HELPSS.

Procedure. A multiple baseline design with multiple probes was used to assess the effectiveness of training across behaviors. The dependent behavior was the syntactic structure of the subject's verbal response to the HELPSS stimuli, while the independent behavior was the training procedure depicted in Figure 2.

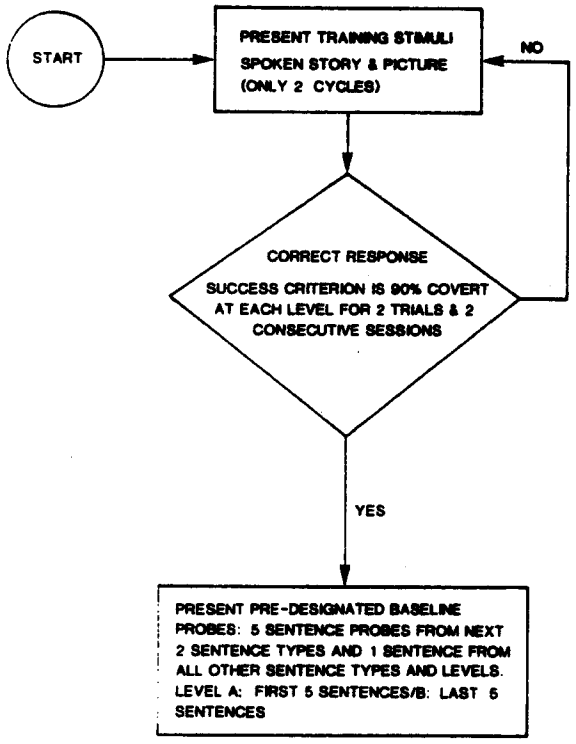

Figure 2. Schematic for treatment procedure - HELPSS.

Instructions for the picture and story presentation were administered in a manner similar to that described in the HELPSS manual. Each subject was administered a sample of the sentence types at each level of complexity, A and B. Sentence types were administered in the hierarchical order presented in the manual.

1. Imperative Intransitive
2. Imperative Transitive
3. WH - Interrogative
4. Declarative Transitive

5. Declarative Intransitive
6. Comparative
7. Passive
8. Yes - No Question
9. Direct and Indirect Object
10. Embedded Sentences
11. Future

The first five picture-sentence items from each sentence type were designated as generalization probes for Level A and the last five picture-sentence items were used as generalization probes for Level B. Scoring of each of the subjects' responses was done according to the system described in the manual (Table 2). As I mentioned earlier the scoring system offered by the manual was not adequate during baseline. However, it was not clear whether the system would be sufficient during the training sessions.

Table 2. HELPSS Scoring System.

Response Description	Score
Fully Correct Response	1.0
Incomplete or Incorrect Response	0.0
Self Corrected Response	0.5

Training. Training proceded as depicted in Figure 3. Two blocks of ten sentences each were presented during each session for a total of 20 sentences. Criterion for moving on to the next level of complexity was two consecutive sessions at 80% accuracy or better per session. During presentation of Level A sentences the Examiner produced a 1 sec pause preceding the target response coupled with a rising intonation pattern. Following each accurate response the Examiner praised the subject. When an error was produced the Examiner said, "No, say this.....," and the correct model was presented. If the subject produced the model accurately he was praised and the next sentence presented. If the subject did not respond accurately to the first model the correct model was presented a second time. Regardless of the subject's response accuracy to this model the next sentence was presented. This training took place in the context of other treatment tasks, so this was not the only treatment being administered during a session. However, the training was clearly separated from other tasks during the session.

Probes of Untrained Sentences. After the subject successfully completed training at Level A and B of each sentence type, all previously trained sentence types were probed and probes were administered on the next two sentence types to be trained. During the administration of the probes neither the pause or the exaggerated rising intonation pattern were used.

Interexaminer Reliability. These data were collected by two different Examiners. Reliability was assessed during less than one percent of the training sessions, and while the reliability was 100 percent for point-to-point reliability, the sample is not as strong as one might desire.

RESULTS

The first question asked was, "Is the training procedure effective in producing accurate grammatical verbal responses to the training stimuli?"

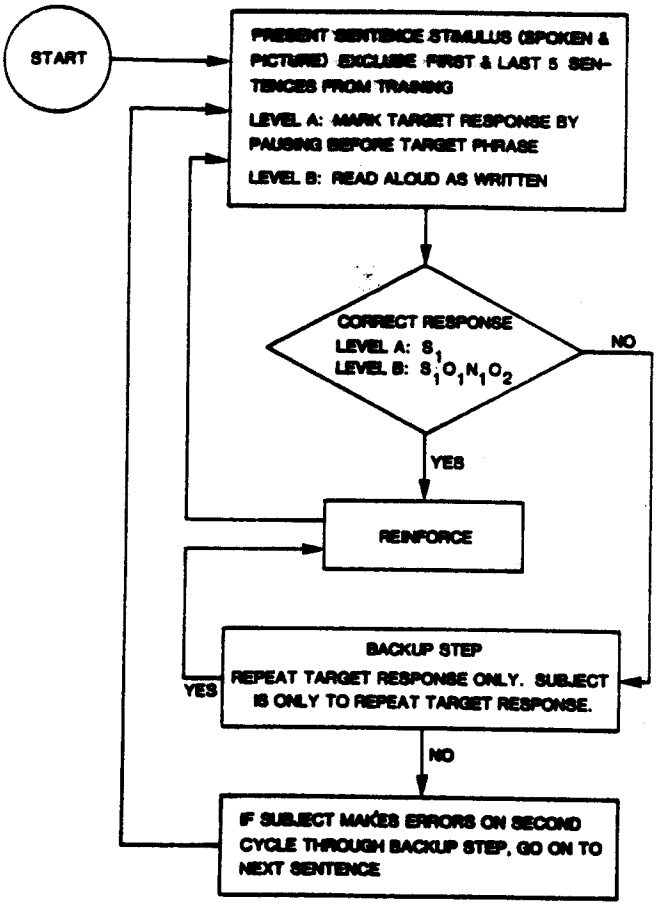

Figure 3. Schematic for training session - HELPSS.

Table 3 summarizes the performance of the subjects. Both reached training criterion within both levels of complexity. Level A took fewer trials to reach criterion than Level B did. However, for two of the subjects, Level A required more trials to reach criterion than Level B did. Figure 4 graphically displays a segment of subject 2's performance across behaviors and sessions. The training data have been collapsed. Each data point represents two training sessions. The probe data have not been collapsed. What conclusions can be drawn from these data? First, training was effective for each level and sentence type. Second, there was little generalization across behaviors preceding training. Third, there was little generalization to untrained sentences following training.

Table 3. Results of training.

	Mean Number of Training Trials to Criterion		Mean Percentage of Correct Responses To Untrained Stimuli			
	Level A	Level B	Level A		Level B	
			Pre-training	Post-training	Pre-training	Post-training
Subject						
1	4.5	3.7	32%	59%	1%	26%
2	3.1	2.1	6%	27%	1%	5%
3	2.2	7.4	94%	98%	15%	16%

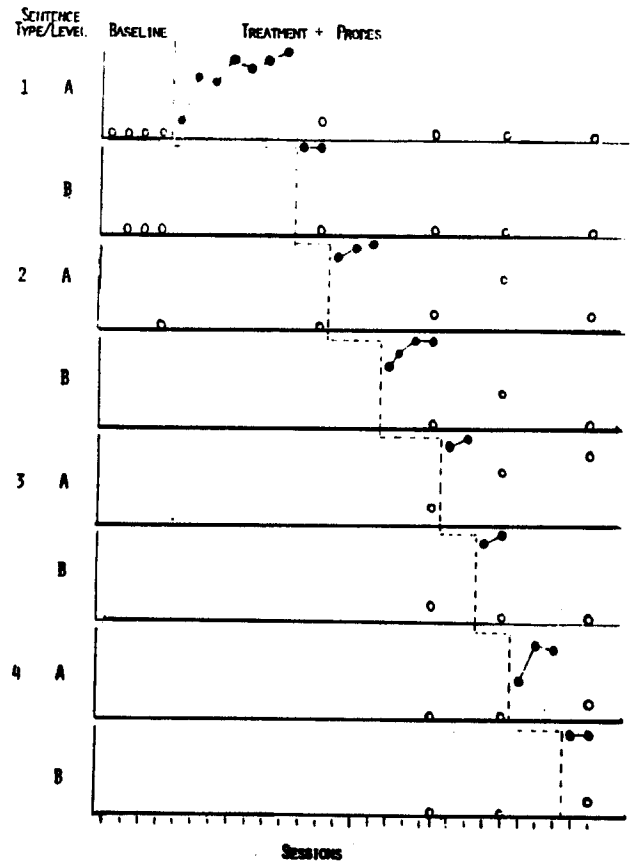

Figure 4. Performance of Subject 2 across behaviors and sessions.

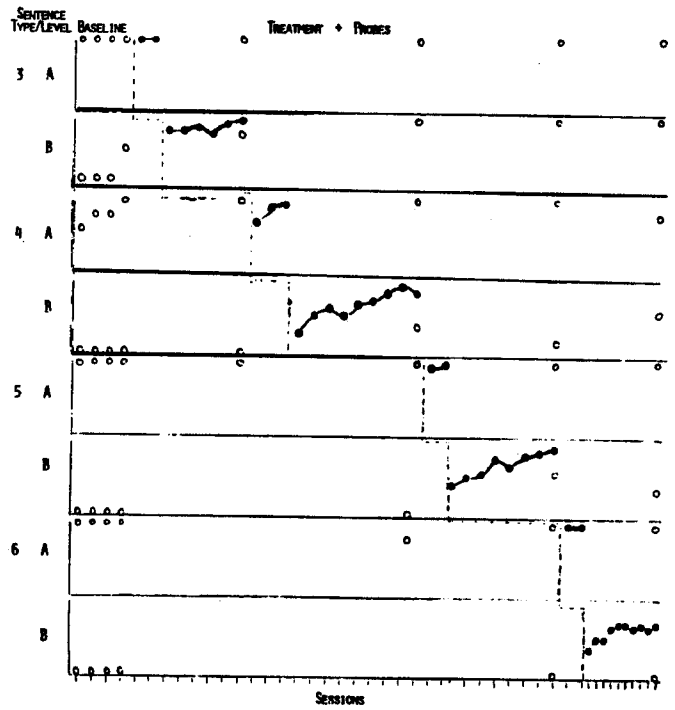

Figure 5. Performance of Subject 3 across behaviors and sessions.

Figure 5 shows a segment of the third subject's performance across behaviors and sessions. These data have not been collapsed. The performance is similar to the previous subject with the exception that he did not meet criterion on sentence type 6, Level B. While it appears that Level B is more complex and thus more difficult, the data suggest the contrary. That is, on the average there was no difference between the two levels in terms of the mean number of trials to criterion. Therefore, the difficulty of Level B must be considered in the context of the influence of training on Level A.

The second question asked was, "Is the training procedure effective in producing accurate grammatical verbal responses to the untrained stimuli?" As you saw in the individual data graphs and the group data presented in Table 3 there is generally an increase in the percentage of probe sentences responded to correctly following training but the differences are really very small. Since the number of probe sentences varied across probe assessments, a percentage calculation was used. Therefore, an increase in responding from 32% to 59% could be only a change from two correct to three correct sentences, while an increase from 6% to 27% indicates no change at all. There is no real difference in the nature of the generalization in Level A or B.

Samples of probe sentences (Table 4) indicate that there was little generalization of grammatical structure to untrained sentences as shown by the performance of two subjects on the same picture-sentence stimuli during the probe assessment over time.

Table 4. Examples of Subjects' responses to probe sentences over time.

Subject	Date	Target Response	Subject Response
1	4/13	He's faster	He running
	8/10		He run
	9/2		He's running
	8/10	The pencils were sharpened	The pencil was dulled
	9/2		The pencil was broken
	10/14		The shapi, the priced
	8/10	Did you buy the paper?	Your quarter my paper
	9/2		Did you 25¢ and paper
	10/14		Did you read the paper?
2	5/27	The report was read.	The papers
	6/26		Composition
	7/20		Teacher was report
	8/17		The brief a note
	11/2		The bullitin was reading
	2/8		Report

SUMMARY

All three subjects were trained to respond to Levels A and B of the sentence types trained. All three subjects showed some generalization to untrained sentences. However this generalization was minimal.

During the treatment of a fourth subject an alternative treatment procedure was developed. The subject in question complained that he could not remember what the target responses were from one session to the next for Level B sentences and pictures. To reduce the demands on recall from one session to the next the following procedure was implemented. During Condition 1 a Level A sentence was presented and then followed by its Level B form. Each of the ten training sentences was trained in this manner. Responses to each of the sentence levels was scored separately. During this condition the subject had little trouble with Level B sentences. However when just Level B sentences were presented during Condition 2 he initially had difficulty, which improved with training over seven sessions. Each session ended with the presentation of Condition 1. This performance suggests a strong memory component to responding accurately to Level B sentences. Furthermore, this procedure of stimulus presentation provides intrasubject replication within and across sessions which helps to isolate the memory component of the HELPSS task. This procedure may also increase the probability of improved generalization, since the procedure appears to train to generalization.

Table 5. Alternative training procedure. Percentage correct responses to combined presentation of Levels A and B, B only, and Levels A and B again.

Session	A and B Condition 1	B Condition 2	A and B Condition 3
1	90 80	10	90 90
2	100 100	40	100 90
3	90 90	70	100 100
4	100 100	70	90 80
5	90 90	90	90 100
6	90 100	80	100 100
7	100 90	100	100 100

REFERENCES

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

Salvatore, A., Holtzapple, P., Trunzo, M.J., Graham, B.A. Investigation of the Sentence Hierarchy of the Helm Elicited Language Program for Syntax Stimulation. In R.H. Brookshire (Ed.), Clinical Aphasiology: Conference Proceedings, 1983. Minneapolis, MN: BRK Publishers, 1983.

DISCUSSION

Q: Maybe if the criterion was three trials or more you might have seen better generalization. Perhaps the behaviors weren't fluent enough, so to speak, before they were probed for generalization. I know we trained three trials at 80% response accuracy but they didn't always meet criteria on the generalization probes.

A: Of course the possibility exists that three trials may be a more successful criterion. That is an empirical question. What we need is more data. It should be clear, however, that replication is impossible with the procedures provided in the manual. Hopefully the procedural description provided in this paper will allow for replication and then, and only then, can we begin to determine the effectiveness of any of the procedural variables.

Q: Just a comment about the HELPSS since it's come up a bunch this session. We did that study two years ago and I was interested in Pat's study where generalization really didn't happen in everyday language, but I think you can take HELPSS as a materials package. But use a different hierarchy because the hierarchy never worked for us. Somehow the second sentence type was always the toughest and you lost the patient before you got to what was easiest for him or her. It has been a good materials package and we do take it into contextual situations and it really works well.

A: The contention has been that the HELPSS is an effective treatment package. The research in 1983 (Salvatore, et al.) demonstrated that the hierarchy was not substantiated, nor was the notion that a pretest was sufficient to measure the effectiveness of the package. The paper today indicates that a reasonable procedure, one very similar to other treatment procedures in the literature, was not effective in generalizing to sentences very similar to the training stimuli. One would expect generalization to sentence-pictures very similar to the training items before one could reasonably expect generalization to vastly different contexts like spontaneous speech. Before a patient can be expected to use these forms in context one would expect that the patient could demonstrate the presence of the behavior in his or her verbal repertoire in situations with rather straightforward response contingencies and simple antecendent conditions. These subjects did not. If the suggestion is that the package is more effective in a "loose training" procedure, then the data need to be gathered before any suggestions are offered.