Syntax Stimulation Revisited: An Analysis of Generalization of Treatment Effects

Ruth B. Fink
Moss Rehabilitation Research Institute, Philadelphia, PA

Myrna F. Schwartz
Elizabeth Rochon
Moss Rehabilitation Research Institute, Philadelphia, PA
Temple University, Philadelphia, PA

Jessica L. Myers
Moss Rehabilitation Research Institute, Philadelphia, PA

Gail Simon Socolof
Winifred Masterson Burke Medical Research Institute, White Plains, NY

Ruth Bluestone
Spaulding Rehabilitation Hospital, Boston, MA

A multiple-probes variant of the multiple-baseline across-behaviors design was used to study the extent of generalization associated with syntax stimulation (SS) training. To assess whether acquisition of specific structures generalized across tasks, we used a specially designed sentence elicitation probe, Picture Description with Structure Modeling (PDSM; Fink et al., 1994). To assess whether training enhanced morphosyntactic production in connected speech, we used the coding scheme developed by Saffran, Berndt, and Schwartz (1989). Four subjects with chronic nonfluent aphasia were trained to produce active, passive, and embedded sentences using materials and procedures from the Helm Elicited Language Program for Syntax Stimulation (Helm-Estabrooks, 1981). These sentence structures were trained in successive phases with generalization probes administered before and after each phase. Three subjects with aphasia served as controls. Strong within-task generalization was observed and, in contrast to previous studies, generalization to the novel sentence elicitation task (PDSM). SS training did not yield measurable gains in narrative production.

Psycholinguistic studies of agrammatism have motivated two quite different approaches to treating this condition. "Syntax stimulation" (SS), commercially distributed as the Helm Elicited Language Program for Sentence Stimulation (HELPSS), aims to facilitate production of morphosyntactic structures using a story completion task with multiple exemplars (Helm-Estabrooks, 1981). "Mapping therapy" (MT) aims to improve the processing of verbs and verb-argument structures. MT protocols vary across studies, but most use receptive or metalinguistic judgment tasks to focus attention on the semantic relations encoded in canonical and noncanonical word orders (e.g., Byng, 1988; Jones, 1986; Schwartz, Saffran, Fink, Myers, & Martin, 1994).

Both SS and MT have been the subject of a small number of efficacy studies. Under a multicentered NIH grant, we are comparing the effectiveness of these two approaches using comparable experimental designs and a uniform set of generalization measures. Here, we report preliminary results bearing on syntax stimulation.

A key concern of the study is how well sentence production training generalizes to materials and tasks different from those used in training. The prior evidence on syntax stimulation is equivocal. On the one hand, there is good evidence that a program of syntax stimulation yields measurable improvement on formal tests and complex picture description (Doyle, Goldstein, & Bourgeous, 1987; Helm-Estabrooks, Fitzpatrick, & Barresi, 1981; Helm-Estabrooks & Ramsburger, 1986). Doyle et al.'s (1987) experimental study also produced evidence of generalization to untrained exemplars of trained sentence types. However, Doyle and colleagues found only limited evidence of generalization across stimulus conditions (hereafter, *across-task generalization*). This is difficult to interpret because the authors did not examine whether the generalization probes reliably elicited the targeted sentences in a control sample. Other experimental studies have used prompted interviews, open-ended conversation, or complex picture descriptions to examine across-task generalization (Kearns & Salmon, 1984; Thompson & McReynolds, 1986; Wambaugh & Thompson, 1989). Such tasks are problematic, first, because there is either no requirement or limited opportunity to use a specific sentence type (see Wambaugh & Thompson, 1989) and, second, because the cognitive and linguistic demands of the task differ considerably from the training tasks.

The present study sought to overcome these limitations by using a new sentence-elicitation procedure specifically developed to measure across-task generalization associated with SS and MT training (Fink et al., 1994). Prior testing with this Picture Description with Structure Modeling

(PDSM) task has shown that it reliably elicits the target structures from subjects with aphasia and non-brain-damaged control subjects, and that it is sensitive to grammatical complexity and to grammatical deficits in persons with aphasia.

A general finding in sentence production treatment studies is that generalization to new exemplars closely parallels training, that is, improvement on exemplars of sentence type B occurs following training on B but not A (for review, Thompson, 1989; and for evidence from an SS treatment study, Doyle et al., 1987). In the present SS treatment study, we sought evidence of across-structure generalization in a within-task (SS) and across-task (PDSM) probe.

Finally, we examined generalization to the more remote, less constrained task of story narration. Saffran, Berndt, & Schwartz's (1989) Quantitative Production Analysis (QPA) yields counts of several structural and morphological elements in the speech samples generated during narration of a familiar tale, for example, the Cinderella story. This Cinderella/QPA analysis makes possible a more detailed assessment of the benefits of sentence production treatment than was possible in earlier studies. For example, SS focuses on morphosyntactic production, whereas MT focuses on verb-argument relations. It might be expected, then, that SS will facilitate the production of grammatical morphemes and MT the production of verbs and more complex argument structures. These predictions for MT have received some confirmation in published studies (Byng & Black, 1989; Schwartz et al., 1994). Here, we tested the prediction for SS.

To summarize the specific questions posed in this study:

1. Does syntax stimulation (SS) training promote acquisition of HELPSS Active, Passive, and Embedded sentences in subjects with chronic, nonfluent, agrammatic aphasia?
2. Does SS training generalize to untrained exemplars tested in the SS (story completion) format? Does such *within-task* generalization extend to sentence structures not yet trained?
3. Does training on a given sentence structure improve performance on trained and untrained exemplars tested in the PDSM format? Does such *across-task* generalization extend to exemplars of sentence types not yet trained?
4. Does SS training improve the grammatical organization of subjects' story narratives?
5. Does SS training improve scores on other measures of language processing (e.g., sentence comprehension, naming)?

Methods

Subjects

Seven individuals with chronic nonfluent aphasia who met the inclusion criteria for the HELPSS target population (grammatical disorders with generally moderate to well preserved comprehension; Helm-Estabrooks, 1981) participated in this study. All were right-handed, had suffered a single left-hemisphere CVA, were native speakers of English, and had passed pure-tone audiometric screening at 40 dB (ANSI, 1989). Four were randomly selected to receive the syntax stimulation (SS) protocol. Three served as controls, yoked to a treated subject of similar aphasia type and severity, and education level (see Table 1).

Control subjects were seen for the same number of sessions as their treated counterparts and given all assessment and stimulus generalization probes at the same time intervals. Their "treatment" session consisted of viewing and retelling a short video segment and engaging in informal conversations with the clinician.

Design and Procedures

A multiple-probes variant of the multiple-baseline across-behaviors design (Horner and Baer, 1978) was used. In conducting this experiment, we used the materials and procedures from the HELPSS protocol with some modifications. The number of structures trained was reduced from 11 to 3, and the number of training exemplars was reduced from 20 to 10. The three sentence structures selected for training, *Active*, *Passive*, and *Embedded*, were chosen to represent a range of syntactic difficulty from the HELPSS hierarchy. They were trained in successive phases with baseline-generalization probes administered before and after each phase.

Baseline Probes

Syntax Stimulation (SS) Probe. This within-task probe was carried out using 10 *untrained exemplars* from the HELPSS program for each of the three sentence types undergoing training. The HELPSS elicitation procedure (stories and pictures) was used to elicit the Active sentences, first at Level A and then at Level B.[1] This procedure was repeated for Passive and Embedded structures, successively. The entire 60-item probe (10 exemplars for three structures, each at Levels A and B) was administered in a single session and repeated for two additional sessions. No specific feedback or training was given during these probes.

Only Level B trials were scored. Following Doyle et al., 1987, we coded a response as correct if it contained all the grammatical elements of the target sentence

TABLE 1. Subject characteristics.

							BDAE		BNT
Subject	Gender	Race	Hand	Education (years)	Age (years)	Years Post-CVA	Severity rating	Mean auditory comprehension percentile	Percent correct
HW	M	B	R	12	55	3.8	2.5	86	68
ES[a]	M	W	R	12	51	6.3	2.5	91	58
GRvs	F	B	R	12	52	5.8	2.0	56	35
CB[a]	F	W	R	14	41	7.3	2.0	56	30
StS	M	W	R	14	62	4.0	1.0	86	63
JW[a]	M	B	R	12	58	2.3	1.0	78	58
PJ	F	W	R	12	55	14.0	2.5	94	NT

[a] Yoked aphasic control subjects

---

[1] In the HELPSS program, a two-level elicitation procedure is used: Level A requires the subject to produce a delayed repetition of the target sentence that has been presented in a story format; Level B requires the subject to complete the story with the target sentence without benefit of repetition.

type and was semantically appropriate to the probe question and the stimulus picture. Self corrections made within 30 seconds were accepted as correct.

Picture Description with Structure Modeling (PDSM; Fink et al., 1994). This 60-item picture description task uses modeling and probe questions to elicit 10 exemplars for each of 6 sentence types. As seen in Table 2, three of these are the SS sentence structures trained in this study. The remaining three are taken from the mapping therapy literature. The target sentences for SS-Active, SS-Passive, and SS-Embedded are the same (untrained) exemplars used in the SS probe, but the *pictures* (as well as the eliciting condition) are novel. The MT-Actives (Type 1) and MT-Passives (Type 2) are morphosyntactically similar to the SS Actives and Passives and were thus predicted to show benefit of training.

Trials on the PDSM are blocked by sentence structure, with examples (models) provided at the start of each block as a way of informing the subject as to the type of structure required. On these modeling trials, the examiner presents a picture, asks a probe question (see Table 2), and then models the desired response using the targeted sentence type. After three modeling trials, the test trials are presented. For each test trial, the examiner presents a picture and a probe question but does not provide a model or feedback. Scores are assigned only to unmodeled trials, using the same 1/0 scoring system as in the SS probe.

In keeping with the findings of the validation study of the PDSM (Fink et al., 1994), the PDSM was administered twice prior to treatment—only the second score was counted—and once following the training of each sentence type.[2]

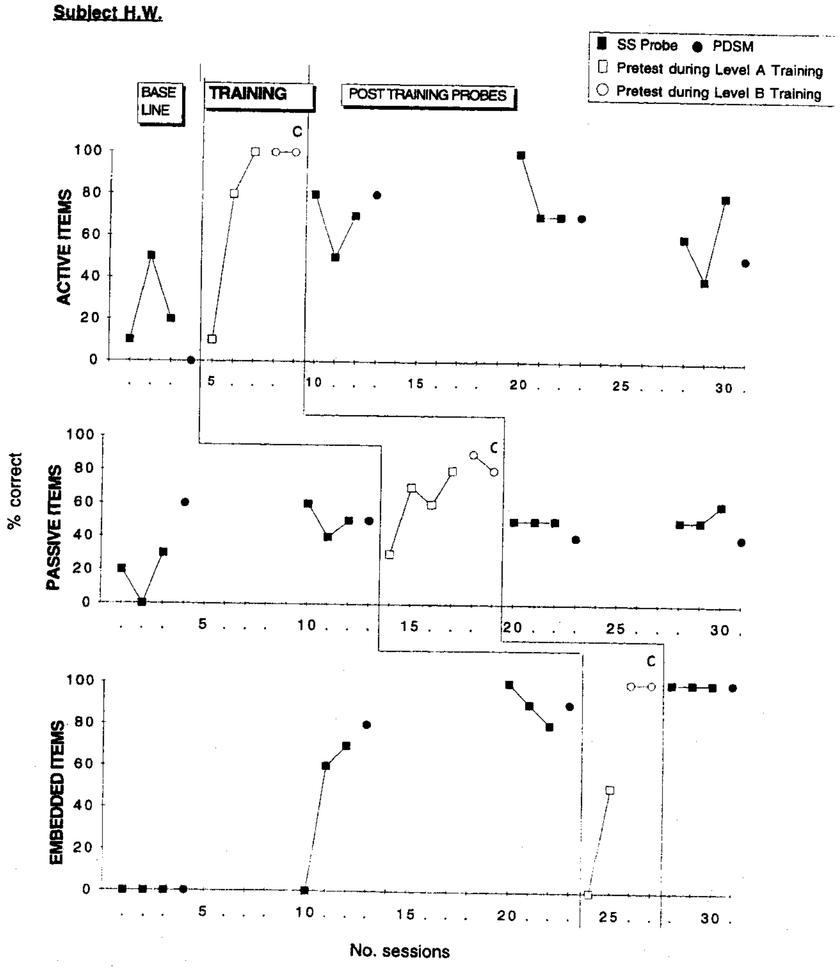
Additional language tests, representing a set of standard dependent measures, were administered to each subject before and immediately after the treatment study. These included the Boston Diagnostic Aphasia Examination (BDAE; Goodglass and Kaplan, 1983); Philadelphia Comprehension Battery (PCB; Saffran, Schwartz, Linebarger, Martin, & Bochetto, 1988); and Boston Naming Test (BNT; Kaplan, Goodglass, and Weintraub, 1983). In addition, a 150-word narrative speech

TABLE 2. PDSM sentence types and probe questions.

Sentence Type	Probe Question	Response
SS Actives	What does he do?	He plays baseball.
SS Passives	What happened to the city?	The city was destroyed.
SS Embedded	What did his mother want?	She wanted him to be happy.
MT Actives (Type 1)[a]	What is the policeman doing?	The policeman is chasing the robber.
MT Passives (Type 2)[a]	What happened to the policeman?	The policeman was shot by the robber.
MT Object Cleft (Type 3)[a]	Who is it that she kicked?	It is the cowboy that the teacher kicked.

[a] Mapping therapy (MT) sentences were chosen to represent a range of syntactic complexity comparable to that of the syntax stimulation (SS) sentences, and are never trained in this experiment.

FIGURE 1. HW's performance (percentage correct) on trained and untrained exemplars across the three training phases (Active, Passive, and Embedded). Daily pretest scores represent performance on the trained Level B exemplars during each treatment session; baseline and post-training SS and PDSM probes represent performance on untrained exemplars of the trained structures. *C* indicates criterion was met at Level B training.



---

[2] In the validation study, the PDSM was administered to subjects with aphasia and to control subjects with no brain damage three times. Both groups demonstrated comparable increases in scores from the first to the second administration with no further change in scores from the second to the third administration. The change from #1 to #2 was interpreted as learning to task. In our current study, we administered the PDSM twice prior to treatment, and used the score from the second administration as baseline.

sample was elicited by asking the patient to tell the Cinderella story (or another familiar fairy tale) after looking through a wordless picture book. The samples were tape recorded and then analyzed using the scheme for quantifying syntactic and morphological aspects of aphasic speech developed by Saffran and colleagues (1989).

Treatment

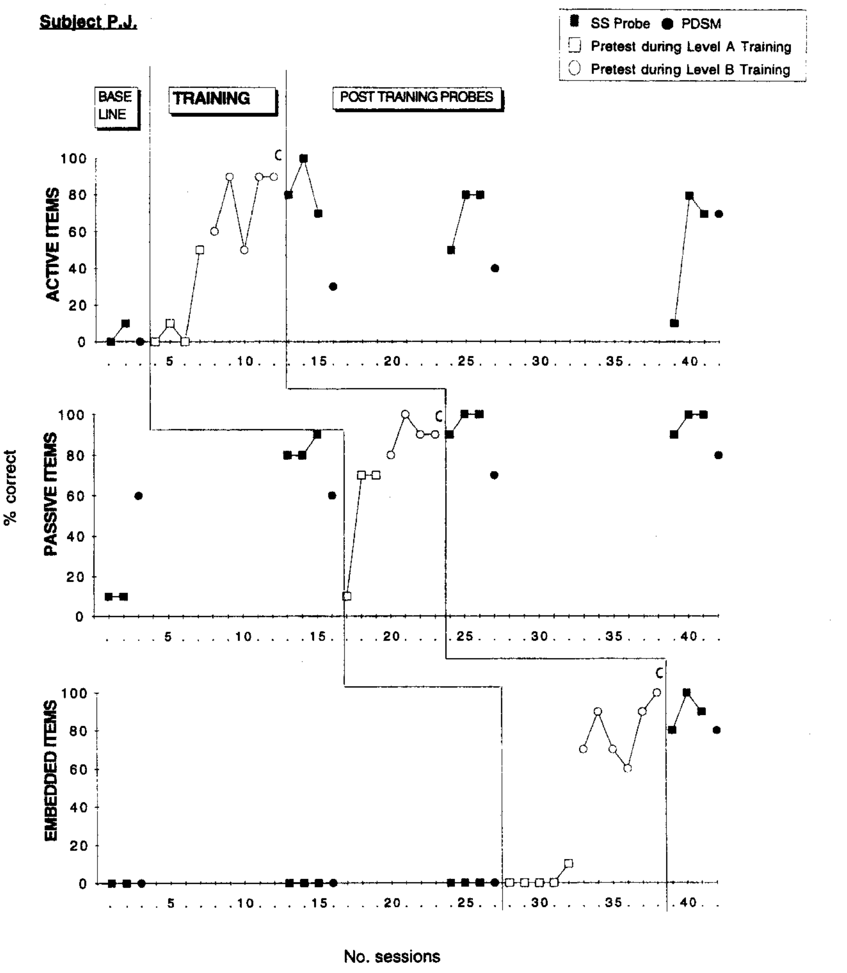
Subjects were seen for treatment three times a week. Each treatment session comprised one pretest block and three training blocks. The 10 exemplars trained in each phase constitute a block.

Training consisted of facilitation of the target exemplars through the HELPSS procedure, first at Level A and then at Level B. Correct responses were reinforced verbally, and incorrect responses were shaped through modeling and forward chaining. Criterion for moving from Level A to Level B was 8 out of 10 correct on three training blocks for two consecutive sessions. Criterion for successful training of each structure (i.e., moving to the next phase) was 8 out of 10 correct on three training blocks (conducted at Level B) for two consecutive sessions plus 8 out of 10 on the pretests given on those days. The order of training phases was Actives, followed by Passives and Embedded.

Pretests elicited the target exemplars at Level B. No feedback of any kind was given. In Figures 1–4, the session-by-session pretest scores are used as an index of acquisition. The same 1/0 system used in SS-probes was used to score pretests and training trials.

To explicitly encourage carryover to more natural settings (Kearns & Salmon, 1984; Thompson, 1989), we concluded each session by having the subject view a 10-minute edited movie and retell the story. Feedback was restricted to general encouragement.

FIGURE 2. PJ's performance (percentage correct) on trained and untrained exemplars across the three training phases (Active, Passive, and Embedded). Daily pretest scores represent performance on the trained Level B exemplars during each treatment session; baseline and post-training SS and PDSM probes represent performance on untrained exemplars of the trained structures. *C* indicates criterion was met at Level B training.



Generalization Probes

At the end of each phase, the SS and PDSM probes were administered. As in baseline testing, the SS probes were given three times, in three consecutive sessions; the PDSM was given once, in a fourth session. Limited training continued during these probe sessions (Koegel & Rincover, 1977).

Reliability

Twenty percent of the pretests, SS probes, and PDSMs were scored by a second scorer, independently (using tapes and loose transcriptions). The percentage of agreement was as follows: pretests: 100%; SS probes: 99%; PDSM: 99%.

Results

Acquisition

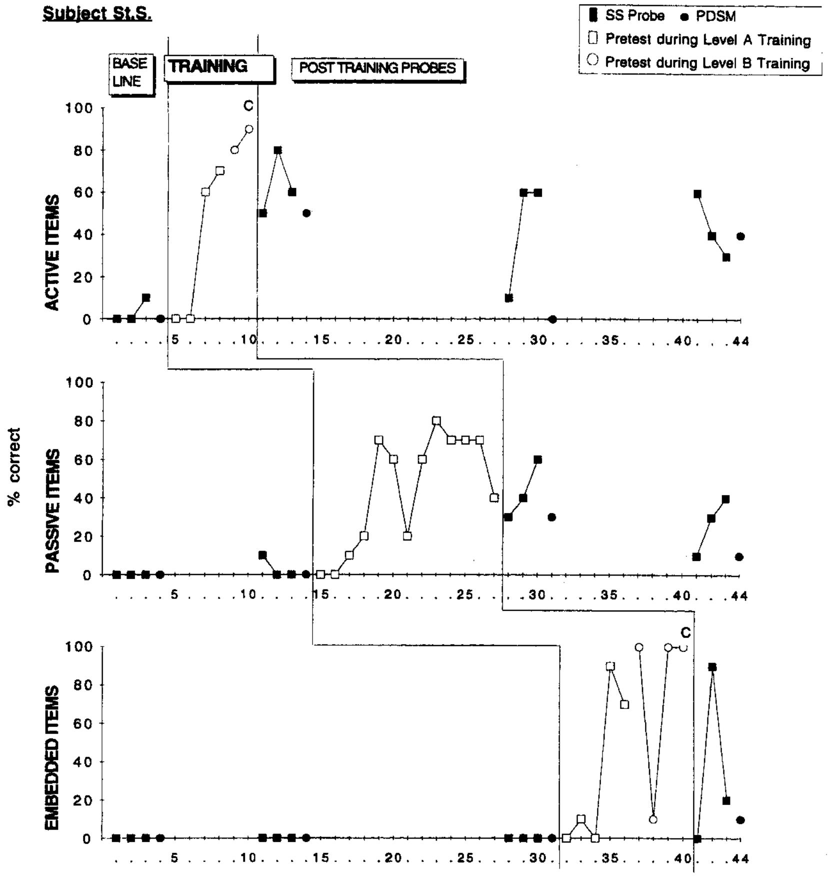
Figures 1–4 chart each subject's performance across sessions. Although baseline performance was variable in some cases (and, in the case of GRvs' Actives, steadily rising), training was initiated after a maximum of three data points had been collected because the similarity of the SS probes to the training task raised the possibility that the probes themselves might constitute a form of training. Acquisition can be assessed by comparing performance on the pretests to the preceding baseline probes. For three of the four subjects (HW, PJ, StS), pretest scores are low early in training, and increase markedly during training. Except for StS on Passives, these three met the acquisition criteria for all training phases.

Within-Task Generalization (SS Probe)

Generalization to new exemplars of the trained structure is reflected in the increase in SS-probe scores immediately following training. The three subjects who showed strong acquisition also showed good generalization on the SS probe.

Across-structure generalization is evident in Figures 1 (HW) and 2 (PJ). HW's improvement following Active training was not limited to Active sentences; Passive and Embedded sentences improved as well following Active training. Similarly, for PJ,

FIGURE 3. StS's performance (percentage correct) on trained and untrained exemplars across the three training phases (Active, Passive, and Embedded). Daily pretest scores represent performance on the trained Level B exemplars during each treatment session; baseline and post-training SS and PDSM probes represent performance on untrained exemplars of the trained structures. *C* indicates criterion was met at Level B training.



Passive production showed the same marked increase as Active production following Active training.

Across-Task Generalization (PDSM)

Generalization to exemplars of the trained structure is evident in the increase in PDSM score immediately following training. Comparison of SS and PDSM probe scores shows that this increase closely paralleled the SS probe. For example, for subject StS, post-training SS-probe scores increased for Actives, Passives, and Embedded sentences; and similar, though less impressive gains were evident on his PDSM probe. Subject PJ showed an especially strong effect for Embedded sentences both on the SS probe and on the PDSM.

Across-structure generalization was also evident on the PDSM, and again, the parallels with the SS probe were striking (see especially Figure 1, Subject HW). Interestingly, the evidence for across-structure generalization was limited to the SS structures and did not extend to the MT structures (Type 1, Type 2, Type 3). We will comment further on this finding in the Discussion.

In contrast, the untreated control subjects with aphasia showed considerably less improvement and more instability across repeated administrations of the PDSM.

Generalization to Other Language Measures

Pre- and post-score changes on the Cinderella narrative for the treated group are small and variable on all measures, and not different from those of the control subjects.

On the Language Assessment Battery (e.g., BDAE, BNT, PCB) both treated subjects and controls made small and variable changes—in both directions.

Discussion

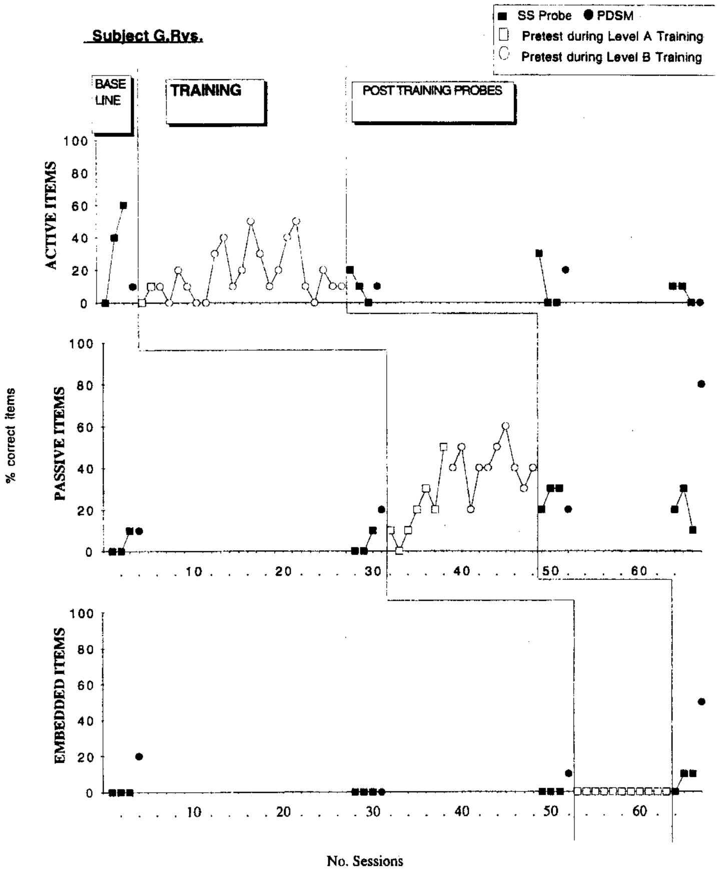
Two of the four treated subjects met our strict acquisition criterion for all three trained structures and a third (StS) met criterion for the Active and Embedded structures. Subject GRvs did not meet criterion for any of three structures, but she did make progress on the Passives. Her performance on the Active sentences was largely hindered by failure to produce a semantically appropriate verb, frequently substituting the verb "does" instead of the main verb. In general, the acquisition data confirm previous reports that the HELPSS protocol is suitable for persons with chronic agrammatic aphasia (Helm-Estabrooks & Ramsberger, 1986; Doyle et al., 1987).

The finding from the SS probe showing that acquisition generalizes to untrained exemplars of trained structures replicates Doyle et al. (1987). That it generalizes to exemplars of *untrained* structures is a new finding. It may be that repeated exposure to the untreated exemplars, which were used in both the SS and PDSM probes, was sufficient to constitute training on these structures or these particular sentences. Some suggestive evidence that the effect is sentence-specific comes from the fact that on the PDSM, the across-structure generalization effect did not extend to the MT-Actives or Passives, despite their structural overlap with the SS-Actives and Passives. If this line of reasoning is correct, the appearance of across-structure generalization on the SS and PDSM probes is illusory and actually constitutes evidence of sentence- or structure-specific acquisition brought about by stimulation in the absence of any feedback.

The evidence for across-task generalization to the PDSM is unambiguous. The procedures and the picture probes used to elicit the sentences differed from SS training and SS probes, and still performance on the PDSM closely tracked the acquisition and generalization patterns obtained there. Thus, with suitable probes (i.e., probes that reliably elicit the structure in contexts that are comparable cognitively and linguistically to the training task), it is possible to show that SS training gains do carry over to novel eliciting contexts.

On the other hand, we did not find that SS training resulted in improved morphosyntactic production on story narration, nor on other measures of language performance. We are currently

FIGURE 4. GRvs' performance (percentage correct) on trained and untrained exemplars across the three training phases (Active, Passive, and Embedded). Daily pretest scores represent performance on the trained Level B exemplars during each treatment session; baseline and post-training SS and PDSM probes represent performance on untrained exemplars of the trained structures.



conducting similar studies with Mapping Therapy to see whether preliminary reports of broader generalization can be replicated (e.g., Schwartz et al., 1994) and whether ordered combinations of MT and SS will produce more extensive gains. Future studies that incorporate generalization training procedures to improve production in progressively more complex tasks are also needed.

Author Note

This study was supported by grant RO1 DC01825 from NIDCD. The authors thank Gwen Dorfman and Paula Sobel for their assistance with data collection and analysis and Cynthia Thompson for her valuable input.

Elizabeth Rochon is now at the Department of Communication Disorders, Baycrest Centre for Geriatric Care.

References

American National Standards Institute. (1989). Specifications for audiometers. (ANSI S3.6-1989). New York: ANSI.

Byng, S. (1988). Sentence processing deficits: Theory and therapy. *Cognitive Neuropsychology, 5*, 629–676.

Byng, S., & Black, M. (1989). Some aspects of sentence production in aphasia. *Aphasiology, 3*, 241-263.

Doyle, P. J., Goldstein, H., & Bourgeois, M. (1987). Experimental analysis of syntax training in Broca's aphasia: A generalization and social validation study. *Journal of Speech and Hearing Disorders, 52*, 143–156.

Fink, R. B., Schwartz, M. F., Rochon, E., Myers, J. L., Socolof, G. S., Bluestone, R., & Sobel, P. R. (1994). *Picture Description with Structure Modeling (PDSM): A procedure for measuring syntactic generalization.* Poster presented at the Academy of Aphasia in Boston, MA.

Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders* (2nd ed.). Philadelphia: Lea & Febiger.

Helm-Estabrooks, N. A. (1981). *Helm Elicited Language Program for Syntax Stimulation (HELPSS).* Austin, TX: Exceptional Resources.

Helm-Estabrooks, N. A., Fitzpatrick, P. M., & Barresi, B. (1981). Response of an agrammatic patient to a syntax stimulation program for aphasia. *Journal of Speech and Hearing Disorders, 47*, 385–389.

Helm-Estabrooks, N. A., & Ramsberger, G. (1986). Treatment of agrammatism in long-term Broca's aphasia. *British Journal of Disorders of Communication, 21*, 39–45.

Horner, R. D., & Baer, D. M. (1978). Multiple-probe technique: A variation of the multiple baseline. *Journal of Applied Behavioral Analysis, 11*, 189–196.

Jones, E. V. (1986). Building the foundations for sentence production in a non-fluent aphasic. *British Journal of Disorders of Communication, 21*, 63–82.

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

Kearns, K. P., & Salmon, S. (1984). An experimental analysis of auxiliary and copula verb generalization in aphasia. *Journal of Speech and Hearing Disorders, 49*, 152–163.

Koegel, R. L., & Rincover, A. (1977). Research on the difference between generalization and maintenance in extra-therapy resonding. *Journal of Applied Behavior Analysis, 10*, 197–206.

Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language, 37*, 440–479.

Saffran, E. M., Schwartz, M. F., Linebarger, M., Martin, N., & Bochetto, P. (1988). *Philadelphia Comprehension Battery.* Unpublished test battery.

Schwartz, M. F., Saffran, E. M., Fink, R. B., Myers, J. L., & Martin, N. (1994). Mapping therapy: A treatment program for agrammatism. *Aphasiology, 8*, 19–54.

Thompson, C. K. (1989). Generalization research in aphasia: A review of the literature. *Clinical Aphasiology, 18*, 195–222.

Thompson, C. K., & McReynolds, L. J. (1986). Wh interrogative production in agrammatic aphasia: An experimental analysis of auditory-visual stimulation and direct production treatment. *Journal of Speech and Hearing Research, 29*, 193–206.

Wambaugh, J. L., & Thompson, C. K. (1989). Training and generalization of agrammatic aphasic adults' wh-interrogative productions. *Journal of Speech and Hearing Disorders, 54*, 509–525.

Contact author: Ruth B. Fink, Moss Rehabilitation Research Institute, 1200 West Tabor Road, Philadelphia, PA 19141.

Key Words: aphasia treatment generalization, agrammatism, syntax stimulation, HELPSS