

PLANNING AND INTERLANGUAGE VARIATION

Graham Crookes

University of Hawaii, Manoa

Having focused previously on attention, cognitively oriented investigations of interlanguage variation and development are turning toward other possible explanatory variables, such as planning. The present study reports on an experiment in which two groups of 20 Japanese learners of English as a second language performed two monologic production tasks with and without time for planning. It was found that providing learners with time to plan their utterances results in interlanguage productions which are more complex in the short run.

The development of a theory (or theories) of interlanguage (IL) variation and change is of central concern to anyone investigating second language (SL) learning. Theories which relate IL variation to cognitive processes capable of manipulation or change are of particular importance, since they may be directly applicable to the learning/teaching situation. Those cognitive processes which have control over others (i.e., metacognitive processes) have been identified as particularly important, and some authorities (e.g., Calfee, 1981) see them as divided into those concerned with planning processes and those concerned with monitoring their operation. The most important cognitively based IL theory concerned with the latter group has been that of Tarone (e.g., 1982, 1983), which has considered the allocation of attentional resources in the carrying out of the cognitive processes involved in SL generation. It has, however, met with some serious problems (Crookes, 1988; Parrish & Tarone, 1986; Rampton, 1987; Tarone, 1987). The most prominent conceptualization of a monitor in SLA theory—that of Krashen (e.g., 1977)—has also encountered some serious criticism (e.g., Gregg, 1984), and Krashen has downgraded its importance so that it now plays almost no part in his work (Krashen, 1982). Accordingly, some researchers have begun to look at other possible causes of IL variation in the cognitive domain, such as the other major category of metacognitive process mentioned earlier—planning. In this article, I will only address the matter of planning: first, briefly outlining some of the previous work in this area, and second, presenting an exploratory study of the effects of planning on the complexity and accuracy of IL.

PREVIOUS RESEARCH ON PLANNING

The idea of planning has been utilized within psychology at least since the publication of *Plans and the structure of behavior* (Miller, Galanter, & Pribram, 1960), particularly in formal modeling of cognitive operations (e.g., Cohen & Perrault, 1979; Hayes-Roth & Hayes-Roth, 1979). In addition, the concept of planning has been widely used in models of L1 language production. Foss and Hakes's summary (1978) is representative:

Sentence production includes the formulation of an idea that initiates an act of speaking and the choice of an appropriate linguistic framework into which to cast it. These, what we might term the "planning" aspects of production, include such things as finding appropriate lexical items to use and arranging them in a suitable semantic and syntactic framework. (p. 170)

On the basis of research in this area concerning the role of the semantic system in the production of extended utterances, Butterworth (1980b) states that "plans for up to about 12 clauses have been reported (Butterworth [1980a], Beattie [1980])" (p. 456). These are semantic plans, that is, ones in which the syntactic and lexical elements have not been explicitly selected. Although debate in this area is far from settled, a number of recent studies using pausal and gaze data also suggest that some planning units are probably "suprasentential in scope and semantic in nature" (Beattie, 1980, p. 81; Holmes, 1984).

Studies of planning carried out within linguistics consist primarily of a small group of papers by Ochs and colleagues (Ochs Keenan & Bennett, 1977). In these papers, a distinction between planned and unplanned language has been made (Ochs, 1979): "1. Unplanned discourse is discourse that lacks forethought and organizational preparation. 2. Planned discourse is discourse that has been thought out and organized (designed) prior to its expression" (p. 55). These investigations of the topic are based on spontaneous oral narratives and written versions of the same narrative by the same individual. However, as Tannen (1982) has observed, the modality confound in this work prevents clear interpretation of the results. A small-scale study which avoided this problem was conducted by Danielewicz (1984), who compared samples of two adults' unplanned spoken language taken from "dinner table conversations" with planned spoken language ("class lectures or prepared talks"). She noted global and specific discourse differences, the former concerning, for example, the way evidence is used to build an argument, and the latter relating to measures of productivity and complexity (words, dependent clauses, coordinate clauses, nominalizations, attributive adjectives, and participles per idea unit), most of which favored the planned condition.

L2 WORK: SECOND LANGUAGE PRODUCTION

By comparison with L1 studies, little attention has been given to planning by researchers concerned with L2 production models, though it is taken as an uncontroversial assumption in many discussions (e.g., Faerch & Kasper, 1983), and Hulstijn and Hulstijn (1984) also take the concept for granted in their work:

The speech production process consists of the conceptualization of a message, the planning of an utterance, and the articulation of the planned utterance. These three processes take place in an incremental and interactive way. . . . Planning involves the activation and retrieval of knowledge about linguistic forms and their meanings, stored in the speaker's memory. It has been suggested that there are several stages in the planning and execution phases of speech production, during which speakers review their utterance plan and may or may not decide to change it. (p. 24)

Much of the small amount of work done on SL production has been collected in Dechert and Raupach (1980) and Dechert, Möhle, and Raupach (1984). These reports of a variety of small studies support the idea that some differences in ILs are attributable to the process of planning, on the basis of data concerning repetitions, corrections, pausal and other temporal measures, and so forth, under various conditions and tasks. The methodology of these studies is purely descriptive, as befits initial exploratory work. Since in addition they are not oriented towards learning, their utility for the investigation of the role of planning in IL development is limited, though they certainly cannot be ignored.

PLANNING AND SL DEVELOPMENT

An implied advocacy of using planning consciously in SL learning has been in existence in second language literature at least since the "good language learner" studies of the 1970s (e.g., Naiman, Fröhlich, Stern, & Todesco, 1978) and has also surfaced in such self-help guides as Rubin and Thompson (1982). One of the few pieces of empirical research on planning in this context is that by O'Malley, Chamot, Stewner-Manzanares, Kupper, and Rosso (1985). They used "functional planning," which "involves having the learner analyze the requirements of a communication task to determine if he or she has the language skills necessary to fulfill those requirements . . . and then proceed to learn new language as required for the task" (p. 573). Since measures of success on the language tasks used were global and the use of functional planning was deliberately grouped with use of other strategies, this investigation unfortunately cannot provide evidence about the specific effectiveness of planning in language production or learning.

The most recent SL work to consider planning is that of Ellis (1985), who finds a very close parallel between Tarone's formal/informal distinction and Ochs's planned/unplanned dichotomy. In discussing SL development, his position is that "one way in which SLA can proceed . . . is by forms which are initially part of the learner's careful style to spread to his vernacular style" (p. 94).

In recent experimental work, Ellis (1987) looked at the differences between adult ESL learners' use of past tense marking in planned and unplanned speech. His subjects performed three tasks. First, they wrote a composition based on pictures illustrating a story, set up to induce use of the past tense (Task 1). Then, they recorded two oral versions of the story, the second of which was analyzed (Task 2). Finally, a second set of pictures was presented for 2 minutes, and subjects recorded an oral version of a story based on the pictures (Task 3). Ellis (1987, p. 8) found that past tense verbs show "a decrease in accuracy from Task 1 to Task 3. For the irregular past

tense, accuracy levels remained more or less constant . . . for the past copula the accuracy levels on Tasks 1 and 2 were almost identical, but on Task 3 they were markedly lower." These results are in line with earlier investigations, such as that of Danielewicz (1984). However, they must be interpreted with caution, as there are conceptual and methodological problems. With regard to the former, Ellis refers to differences in use of the syntactic forms under consideration as "style-shifting," while at the same time making it quite clear that the differences have been induced by providing subjects with greater or lesser opportunities to plan; the latter include a method of counting correct occurrences which ensures that occasions of correct use of the past tense will not generally be counted if they occur after an initial incorrect use. (See Preston, 1989, p. 287, fn. 6.)¹

In sum, work done in this area so far has tended to indicate that for non-native speakers (NNSs), planned speech differs from unplanned in the same respects as it does for native speakers (NSs). In addition, there seem to be differences (such as suppliance of morphology) which do not appear in comparisons of NSs' planned and unplanned language, which could be implicated in IL change toward target forms. Consequently, the following question can be posed: Does inducing NNSs to plan their speech result in their producing interlanguage which differs from that produced without planning in ways relevant to SL development?

HYPOTHESES

This question can be formalized in a variety of ways. In particular, one could look for change in the use of elements of developmental sequences, overall measures of variability, and broad measures of IL development. The general hypothesis here is that planned speech will show more evidence of development than unplanned speech in a variety of respects. This is broken down into a series of specific directional hypotheses.

Hypothesis 1: Within any specific grammatical domain, there will be a greater ratio of different forms to total possible forms. This relates to the question of variation and utilizes a basic measure of this. If planning results in the use of forms higher in developmental sequences, there may be increased suppliance of those forms at the same time as lower level forms are still in use. Thus, under planned conditions, there might be more variability of IL, in the sense that a greater range of forms is being used.

Hypothesis 2: There will be a greater number of words per utterance. This follows directly from the work of Danielewicz (1984), who found this difference in L1 planned speech, as mentioned earlier.

Hypothesis 3: There will be a greater number of words per error-free T-unit (Hunt, 1966); and

Hypothesis 4: There will be greater use of error-free T-units. (Hypotheses 3 and 4 are derived from the work of Larsen-Freeman [1978, 1983], who has shown that length and proportion of *error-free* T-units in a given passage can be used as rough discriminators of ability levels for SL speech and writing.)

Hypothesis 5: There will be a greater use of complex noun phrases (e.g., adjectival sequences). This concerns differences in the descriptive complexity of planned

speech observed in preliminary work done for the present study. Transcripts appeared to show a greater extent of detailed description under planning. One simple way this aspect of language use can be quantified is in terms of adjective sequences—sequences of adjectives strung together to provide increasingly exact, more complex modifications of the referent identified by a particular noun.

Hypothesis 6: There will be a greater number of subordinate clauses per T-unit. The work of Gaies (1977) indicates that provision of subordinate clauses is a potentially sensitive measure of ESL development. Chaudron (1979) also found this measure indicative of the level of ESL teachers' speech complexity, and its use is further supported by Danielewicz's (1984) findings concerning native speakers.

Hypothesis 7: There will be greater correct use of grammatical functors. The use of morphology is generally recognized as an index of development in the acquisition of English as an SL (Hatch, 1983; Long, 1988).

Hypothesis 8: There will be greater correct use of forms higher in (appropriate) established developmental sequences. As is well known, development of English IL proceeds in some areas via a fairly regular sequence of different manifestations of the same underlying form (e.g., negation). In these areas, increased suppliance of a higher form is indicative of development. (See, e.g., Bailey, Madden, & Krashen, 1974; Cancino, Rosansky, & Schumann, 1975; Hatch, 1983, pp. 89–108, Larsen-Freeman, 1975.)

Hypothesis 9: A greater proportion of T-units will demonstrate forms typical of a higher developmental level. In addition to the classic work mentioned, theoretical work supported by evidence from the acquisition of German as a second language (Meisel, Clahsen, & Pienemann, 1981; Pienemann, 1984; Pienemann & Johnston, 1987) makes it desirable also to consider some more speculative sequences in a related hypothesis.

METHOD

Design

A single-factor repeated measures (RM) design was used, involving two levels of planning time within subjects. Two language production tasks, with 20 subjects per task, were used between subjects. The tasks differed with respect to stimulus materials, though both required the production of monologues. Order of the planning condition was counterbalanced across subjects.

Two different, but equivalent versions of the same tasks were designed, to minimize practice effects from one planning condition to the other. These were counterbalanced across subjects. A distractor task was used between the two task versions to minimize any carry-over effect.

Tasks

Two main tasks were used to eliminate the possibility of results being an artifact of task (Chaudron, 1985). Task 1 was to give a description of how to construct a certain configuration of Lego blocks; Task 2 was to give an explanation concerning the siting

of a set of buildings on a map of a town. In both cases, two equivalently complex versions were developed. Both constructions utilized exactly the same pieces to make two somewhat similar house-like forms. The pieces are made of brightly colored plastic and fit together tightly. Participants were not permitted to take the constructions apart.

The two versions of the map task each involved a simple, stylized sketch map of a town. In both versions the same elements were used within the same basic framework to create problem situations which were equivalent, but sufficiently distinct to constitute different problems.

Operationalization and Levels of Independent Variable

Planning was operationalized in terms of time at two levels. In the *minimal planning* condition, participants were given no time to plan at all, but were instructed to begin their explanation as soon as they had read the instructions; all participants cooperated.

In the *planning* condition, participants were given 10 minutes and told to plan their explanation in terms of words, phrases, and ideas. They were to work independently. They were asked to make written notes on a sheet of paper in English, but were specifically instructed not to attempt to write out in detail everything they would say. They were also told that the paper would be removed at the end of the planning period and that their oral production would be made without it. The reason for having them produce notes was to ensure that they did in fact engage in planning and to have evidence of this. At the same time it was considered most important to avoid the confound of modality which existed in previous work (Ellis, 1987; Ochs, 1979)—hence, the requirements for notes, rather than continuous prose, and the removal of the notes prior to speech. Participants complied with these instructions.

Between each of the two versions of a task, participants completed a detailed questionnaire in Japanese, containing questions which asked for biodata information or concerned their language learning experience. This task was intended principally as a distractor task. After all subjects had taken part, and there had been some initial evaluation of results, they were recontacted and provided with a written explanation of the study and preliminary results, plus the opportunity for further follow-up discussion of the investigation.

Subjects

Subjects were 40 adult non-native speakers of English, all of the same L1 background—Japanese. Having participants from a single L1 background obviates the need to consider possible threats to validity caused by varying L1 transfer effects, culturally conditioned learning styles, and comparative imbalances in command of different modalities across groups with differing sociocultural backgrounds.

Participants were university students with intermediate or advanced levels of spoken English as an SL studying on the Manoa campus of the University of Hawaii (age range: mid-20s to mid-40s; 14 male, 26 female; TOEFL scores 430–650).² Those who agreed to participate were paid \$10 in cash or were offered one hour of tuition

with the experimenter in exchange for their time. This sample was considered adequate on logistical grounds, particularly in the absence of any previous work which could have been used to estimate effect size.

Experimental Procedures and Instructions

Group administration and collection of data (on audiotapes) were done in a language lab. There were 5 or less volunteers per data collection session, resulting in a total of 12 sessions overall. Participants were randomly assigned to treatments (Task 1 or 2, planning first or second) as they entered the lab. They were then reminded in English that their participation was voluntary, that all details of the proceedings would be confidential, that there was no connection with any class or grading procedure on campus, and that if they wished to leave at any time, either then or during the experiment, they would be free to do so. They also were given a written statement to this effect in Japanese.

Participants then received a brief oral explanation in English of the rationale for the experiment, similar to the brief explanation the experimenter gave when they were initially contacted. They were also provided with a detailed set of instructions in Japanese (a translation of English instructions written by the experimenter), along with the actual experimental materials.

Inter-Rater Reliability and Data Analysis

After transcription, measures of inter-rater reliability were obtained on a stratified random sample³ of the discourse produced by participants, for segmentation of the stream of speech into utterances, and for the segmentation of utterances into T-units. Inter-rater reliability (percentage of agreement) for the former was 92%, and for the latter, 90%.

Statistical analysis of data was performed using the statistical package SPSSx, version 2.2 on the University of Hawaii IBM 3081, using the MANOVA program. The multivariate regression approach to repeated measures designs was used (O'Brien & Kaiser, 1985), with directional hypotheses in the univariate tests.

In addition, effect sizes were calculated. Effect size, while "most important" is "the least familiar of the concepts surrounding statistical inference"; it provides an indication of "the degree to which the phenomenon under study is manifested" which is independent of sample size (Cohen, 1969, p. 10). Various specific measures are available, and in this case, the effect size measure η^2 is reported (Cohen, 1973, 1977; see also, Rosenthal & Rosnow, 1984, for an example of its use with repeated measures).

RESULTS

Order Effects

For all measures and hypotheses, the first analysis performed related to the possibility of an effect for order of planning condition. It was necessary to check whether

participants who experienced the planned condition before the unplanned condition performed better than those who experienced the opposite order. Effects for the order of administration of condition on more than a few measures would suggest that there had been a general carry-over effect and would make results much more difficult to interpret. For all but one case, results were non-significant at the .05 level, with the exception being error-free T-units per utterance on Task 2 ($p < .05$, $F = 91.7$, $df = 1/36$). Inspection of the transcripts offers no obvious clues as to why this effect was found on just this task and measure out of the entire group.

For each task, a multivariate analysis of variance was done on the planning factor with the eleven dependent variables which did not apply to acquisition sequences. (The acquisition sequence group was considered too non-homogeneous for this procedure to be appropriate.) The values obtained were: for Task 1, Pillai's = 0.56, $F = 1.04$, $p = .48$; for Task 2, Pillai's = 0.69, $F = 1.84$, $p = .18$. Because of the exploratory nature of the study, individual pairwise differences were also considered between each dependent variable measure on the planning factor.

Main Hypotheses

Table 1 presents a summary of means, standard deviations, and effect sizes on all the measures used. (The latter are reported separately at the end of this section.) With regard to the results on individual hypotheses, two simple test cases were selected for the test of Hypothesis 1—verb phrase and lexis. Hypothesis 1 stated that within a given domain, there would be a significantly greater ratio of different forms to total possible forms. A test of this hypothesis by way of a measure of the range of verb phrase (VP) elements utilized by participants⁴ was not supported (Task 1: $F = 1.12$, $df = 1/19$, n.s.; Task 2: $F = 0.02$, $df = 1/19$, n.s.). For speakers' choice of lexis, however, Table 1 (line 2) shows that the group differences on both tasks favored the planning condition, with the difference on Task 1 being significant ($p < .05$, $F = 3.79$, $df = 1/19$).

Hypothesis 2 predicted a greater number of words would be used under planned conditions, and the test of this hypothesis used a simple measure of complexity of speech produced—words per utterance. Both tasks produced significantly longer utterances under the planned condition (Task 1: $F = 12.88$, $df = 1/19$, $p < .001$; Task 2: $F = 8.39$, $df = 1/19$, $p < .005$) (see Table 1 [line 3]).

Hypothesis 3 predicted a greater length of error-free T-units (in words) under the planning condition. As seen in Table 1 (line 4), mean differences were in the predicted direction on both tasks, but were not significant (Task 1: $F = 2.19$, $df = 1/19$, n.s.; Task 2: $F = 0.91$, $df = 1/19$, n.s.).

Hypothesis 4 stated that there would be greater use of error-free T-units under planning. Table 1 (line 5) shows means and standard deviations for the analysis of data on this measure—no significant differences were observed (Task 1: $F = 1.09$, $df = 1/19$, n.s.; Task 2: $F = 0.02$, $df = 1/19$, n.s.). Caution must be observed in interpreting this result, since as mentioned earlier, an effect for order on Task 2 was observed for this measure.

Hypothesis 5 concerned a simple measure of descriptive language—the use of adjective sequences in noun phrases. The hypothesis predicted greater use of adjectives

Table 1. Summary of results

Measure	Task 1					Task 2				
	+ Planning		- Planning		Effect Size (eta)	+ Planning		- Planning		Effect Size (eta)
	M	SD	M	SD		M	SD	M	SD	
VP range/utterance	0.10	0.04	0.09	0.04	0.29	0.19	0.06	0.19	0.04	0.03
Type/token ratio	3.57	0.92	3.20	0.66	.05	5.40	1.20	5.00	0.98	0.29
Words per utterance	13.65	3.71	11.47	2.40	.05	16.16	3.74	13.38	2.99	.05
Words per error-free T-unit	4.36	3.56	3.41	2.61	0.32	4.24	3.24	3.83	3.13	0.21
Error-free T-units per utterance	0.39	0.24	0.34	0.22	0.24	0.36	0.23	0.36	0.25	0.03
Adjective sequences per utterance	1.44	0.50	1.29	0.44	0.32	0.67	0.36	0.66	0.39	0.01
Words per adjective sequence	3.75	1.62	3.12	1.10	.05	1.69	0.86	1.53	0.83	0.15
Subordinate clauses per T-unit	0.16	0.16	0.11	0.11	.05	0.76	0.48	0.71	0.40	0.15
Subordinate clauses per utterance	0.18	0.19	0.11	0.10	.05	0.86	0.47	0.69	0.47	.05
S-nodes per subordinate clause	1.27	0.31	1.10	0.27	.05	2.00	0.63	1.66	0.54	.05
Words per subordinate clause	1.47	1.73	0.77	0.80	.05	6.36	3.73	5.15	3.15	.06
TLU for plurals	0.72	0.20	0.66	0.21	0.25	(not tested for task 2)				
TLU for the	0.76	0.17	0.68	0.20	.05	0.50	0.25	0.46	0.28	0.23
TLU for a	0.47	0.25	0.42	0.33	0.15	0.61	0.30	0.56	0.39	0.12
Plural concord	0.55	0.25	0.55	0.27	0.00	(not tested for task 2)				
Discourse markers per utterance	0.28	0.08	0.24	0.09	.05	0.45	0.16	0.39	0.12	0.32

tive sequences under planned conditions. This was operationalized in two ways: with regard to adjective sequences per utterance and with regard to words per adjective sequence. Table 1 (lines 6 and 7) show the means and standard deviations for this hypothesis with respect to the two measures used. No significant differences for adjective sequences per utterance were observed (Task 1: $F=2.15$, $df=1/19$; Task 2: $F=0.00$). However, for words per adjective sequence, a significant difference was observed on Task 1 ($F=5.14$, $df=1/19$, $p<.05$), but not on Task 2 ($F=0.45$, $df=1/19$, n.s.).

Hypothesis 6 concerned the complexity of speech produced. It was predicted that there would be a greater number of subordinate clauses per T-unit used under planned conditions. Group differences on the measure were in the predicted direction (see Table 1 [line 8]), but not statistically significant (Task 1: $F=2.85$, $df=1/19$, $p=.054$; Task 2: $F=0.50$, $df=1/19$, n.s.).

The suppliance of subordinate clauses was also investigated in terms of a base unit of utterances. This measure of language complexity (subordinate clauses per utterance) showed a significant difference for planning on Task 2 ($F=4.36$, $df=1/19$, $p<.05$) and on Task 1 ($F=5.28$, $df=1/19$, $p<.05$) (see also Table 1 [line 9]).

An additional measure was used: S-nodes per utterance. This covered all untensed verb forms used in an utterance—it provided a broader measure of complexity than subordinate clauses per utterance. Significant differences were found for both tasks (Task 1: $F=8.33$, $df=1/19$, $p<.005$; Task 2: $F=7.25$, $df=1/19$, $p<.01$) (see also Table 1 [line 10]).

Finally in this area, an additional test of complexity was performed on the length of subordinate clauses. Longer subordinate clauses were used in the planned condition, significant for Task 1 ($F=4.82$, $df=1/19$, $p<.05$), and for Task 2 ($F=2.70$, $df=1/19$, $p=.06$) (see also Table 1 [line 11]).

Hypothesis 7 predicted greater correct suppliance of grammatical functors under planned conditions. It proved somewhat more difficult to test than had been anticipated. English has only a limited amount of bound morphology, and some obvious candidates for testing (such as past tense *-ed*) were almost never utilized by participants in this study on the two tasks used. The most salient feature in the data for this category was plural *-s*, which was called for extensively in Task 1, but was almost non-existent in the discourse on Task 2 (an average of two cases per speaker per session). Accordingly, only the data for Task 1 were analyzed. As can be seen from Table 1 (line 12), the difference in group means for target-like usage (TLU; see Pica, 1983, 1984) favored the planned condition, but was not significant ($F=1.17$, $df=1/19$).

Hypothesis 8 predicted greater use of forms higher in developmental sequences. The test case of *article* was utilized, chosen in particular since Master (1988) has pointed out that acquisition of this item can be established with a relatively small sample of discourse, and because it is a feature of English that is notably problematic for Japanese speakers.⁵ Table 1 (line 13) shows the means and standard deviations for use of *the*, and Table 1 (line 14) shows the results for suppliance of *a* (measured in both cases by way of TLU; Pica, 1983). A significant difference was found favoring the treatment condition for *the* in Task 1 ($F=5.64$, $df=1/19$, $p<.05$), but not in Task 2 ($F=0.86$, $df=1/19$). No significant differences were seen across conditions and tasks

for *a* (Task 1: $F=0.38$, $df=1/19$; Task 2: $F=0.25$, $df=1/19$). In fact, given the nature of the discourse, Task 1 seems to have provided many more possible contexts for use of *the* than Task 2; given the amount of shared information between speaker and purported hearer, there was little need to make use of *a*. The data thus provide partial support for Hypothesis 8.

Hypothesis 9 referred to developmental sequences predicted by the Pienemann-Johnston model (Pienemann, 1984; Pienemann & Johnston, 1987) and predicted that a greater proportion of T-units in planned productions would demonstrate forms of a higher developmental level, with reference to the Pienemann-Johnston model. Given the restrictions of the data collected, the sequence proposed by Pienemann that was most easily investigatable in the present study was for noun affixes: plural *-s*, possessive *-s*, followed by *plural concord* (the use of plural *-s* in agreement with a preceding quantifier, such as *two* or *many*). If the Pienemann-Johnston sequence is correct, and if planning makes a difference to the suppliance of features in this sequence, it might be expected that there would be greater evidence of plural concord in planned productions as a proportion of T-units showing plural concord to those showing regular plurals. As discussed with regard to Hypothesis 8, noun plurals (and contexts for plural concord) were supplied extensively in Task 1, but there was almost no use of them in the discourse of Task 2. (Possessive *-s* was absent in both tasks and so was not included in the test.) This hypothesis was therefore tested only on the data of Task 1. As can be seen from Table 1 (line 15), there was in fact no difference in the suppliance of plural concord across planning conditions ($F=0.00$, $df=1/19$).

Finally, although this was not originally a prediction, inspection of transcripts suggested that Danielewicz's (1984) comments concerning global differences in the organization of discourse for first language planned speech might also apply in second language speech. The number of discourse markers (such as *first*, *second*, *finally*, etc., and *if... then, even though*, etc.) utilized per utterance in planned and unplanned productions was taken as a rough indicator of the level of organization of the discourse. (See, e.g., Rubin, 1982, for application of a similar measure of the logical structure of discourse in child L1 studies.) Table 1 (line 16) shows the means and standard deviations of discourse markers by planning condition and task. Differences favored planning and were significant for Task 1 ($F=3.23$, $p<.05$, $df=1/19$), but not for Task 2 ($F=2.52$, $df=1/19$).

As may be seen from Table 1, effect sizes were calculated and ranged from 0.0 to 0.65. Twelve effect sizes were medium (0.3–0.49), and four were large (≥ 0.5), using the conventional figures of Cohen (1977; see also Rosenthal & Rosnow, 1984, p. 360).⁶

DISCUSSION

This study investigated the nature of IL produced after time for planning, with specific regard to its variability and targetlikeness, by comparison with that produced with essentially no planning time. Results showed that under conditions of planning, NNSs produced a greater variety of lexis (on one task), but not of syntax (in the test case of verb phrase). On both tasks, non-native speakers produced significantly more complex language, as measured in terms of words per utterance, number of subordi-

nate clauses per utterance, and S-nodes per utterance. General measures of accuracy (such as number or length of error-free T-units) did not show significant differences between conditions. Overall, on the twelve measures not directed towards specific developmental sequences as applied to each of the two tasks, all non-statistically significant differences were in the predicted direction, favoring the planned condition (with two cases of no difference on Task 2). For this group of tests, 15 of the 24 effect sizes were greater than 0.3, i.e., of medium or larger size.

Specific measures of accuracy were applied to the use of morphology and to two developmental sequences. It was observed that the use of morphology as tested with reference to the plural marker *-s* did not appear to be improved by the availability of time for planning (although raw group differences were in favor of the planned condition). Within the stages of acquisition of the article, target-like use of the definite article *the* improved when it followed time for planning. Significant differences were not observed under planning for the target-like usage of the usually later acquired *a* (though differences favored planning). Given that some aspects of the article system are acquired late by Japanese (Hakuta, 1978), and that increased target-like usage of a stage part-way up a developmental sequence is usually a precursor to improved usage of higher level units, the result for *the* may be taken as partial evidence that planning allows the NNS increased opportunity to use forms higher in a developmental sequence.

A test of a sequence predicted by the Pienemann-Johnston model concerning noun plurals showed no evidence that development of more target-like production is facilitated by planning in this area, though since the sequence is itself speculative, this is not necessarily evidence against the utility of planning. It is perhaps surprising that results concerning use of morphology were not clearer, given the effects found by Ellis (1987) for increased use of *-ed* with planning. However, besides the possible distorting effects of the written mode used in Ellis's first task, the use of a consistent past context for the written and spoken productions may have sensitized participants in that study to past tense forms. Although use of morphology is generally a measure of SL development for English, formal and informal learners show different patterns in this area, with those learners whose learning has been particularly via instruction oversupplying morphology. Given the fact that all learners in the present study were Japanese, and given the nature of the Japanese education system, the entire group were learners whose primary experience of ESL was through formal instruction. Possibly, as a result, a test of morphology use might be less likely to show differential effects for planning on such a group than on a group of exposure-only learners (cf. Long, 1988; Pica, 1982). A final consideration with regard to the Pienemann-Johnston sequence is simply that the effects of planning on the use of developmental features may show up more clearly with groups of beginners, since many of the Pienemann-Johnston features are largely controlled by intermediate-level learners.

The test of the quality of planned versus unplanned discourse with regard to explicitness of description (use of adjective sequences) showed a difference on one of the two tasks. This was Task 1, the Lego task, where a clear and detailed description of physical items was particularly important (or at least, all participants behaved as if it were a concern for them). In Task 2, the Map task, there was much less need for

detailed description, and so, with hindsight, it seems reasonable that planning would have less effect there. The related test of provision of discourse markers as an indication of the organizational quality of the planned discourse found a statistically significant difference only on Task 1. Direct inspection of transcripts suggests that this may not have been simply the result of an absence of variation in performance across levels of planning, but because the response to planning in terms of discourse organization was different across subjects. Some subjects indeed responded to the availability of time for planning by supplying discourse markers. Subjectively, it appears that others seemed to prefer to provide a more clearly organized account by way of concision and directness. (No attempt was made to quantify this observation.) A few participants seemed to attempt more detailed discussion of reasons why (as in Task 2) a given item should be sited in a specific place, but were not always successful, resulting in some sections of their planned explanation being less well-structured than their unplanned discourse, since they contained more false starts and uncompleted segments of discourse.

The general pattern of these results shows consistent, small- to medium-sized effects in favor of the planned condition and is tentatively taken here as supporting the position that planning is a process that can lead SL learners to produce more developed speech in the short term. Besides considerations pertaining to individual hypotheses, there are a number of factors which may explain the absence of stronger results.

Having participants from a single language group has both positive and negative aspects. Although this allowed the factor of first language transfer to be controlled, it allows the possibility that culture-specific patterns of language use may have weakened experimental effects. It may be the case that spontaneity in second language use is not something commonly found among Japanese, and thus a predisposition towards the use of planning (both co-planning and pre-planning) may have limited the differences which might otherwise have been seen.⁷

The effect of cultural factors may have been exacerbated by the social context of the data collection procedures. Participants produced their recordings in a language lab, which was an unfamiliar environment for them, and the researcher was easily identifiable as an ESL instructor (although not the instructor of any of the students). Although participants were told that the experiment was not a test and had no connection with their grades, the effect of the environment may have overcome the message conveyed by the researcher.

There is also the possibility of a connection between the complexity of language produced and its accuracy. It is unlikely that SL learners who produce more complex speech than they are normally capable of will at the same time maintain a given level of accuracy (or closeness to target norms). As the utterances of a second language learner become longer and more complex, the chances of their being completed without error becomes smaller, other things being equal. Since in the present experiment, effects for complexity were most clear, this may explain why there was comparatively little increase in measures of accuracy.

Finally, there is the matter of task. Referring again to Table 1, it may be seen that on most measures, group means for Task 2 were higher than those for Task 1. That

the two tasks called forth different language in some areas is clear, as in the instance of Task 1 requiring detailed description and specification of the various blocks which made up the construction, which shows up in the use of adjective sequences. In general, there was no need for speakers to explain the directions they were giving in Task 1, whereas in Task 2, almost every speaker explained the course of action he or she was proposing. It would appear that Task 2 called for more complex explanations and thus required more from its participants. The nature of the design did not call for an equating of the two groups, and there is the possibility that the average proficiency of the participants who did Task 2 was simply greater than that of those who performed Task 1. If, however, it is assumed that they were equivalent groups, then it could be argued that perhaps choice of task is more important to quality of speech produced than whether or not participants have time to plan. To this the response must be that the importance of task choice is not in question and has been documented elsewhere (Crookes, 1986). The interesting point is that whatever the effects of planning, in the present case, they appear to apply regardless of task (except in the trivial case where one task involves some forms that another does not call for, in which case, obviously, an effect for planning will not be seen).

Classroom Implications

Some of the most promising general developments in SL classroom methodology are those associated with the terms *communicative approach* (e.g., Littlewood, 1981) and *proficiency-based approach* (e.g., Omaggio, 1986). Although research evidence in favor of these positions is scanty, many of the classroom techniques advocated by proponents of these approaches are compatible with current research-based conceptualizations of SL classroom learning. However, planning may be one area where the fit between defensible current SL pedagogy and research evidence is less satisfactory. In discussing the communicative approach, Brown (1987, p. 213) identifies the use of spontaneous, unrehearsed (i.e., unplanned) language as one of four common characteristics of communicative language teaching. The present study would seem to imply the desirability of investigating the classroom use of some non-spontaneous, planned language as a means of promoting SL development.⁸ Furthermore, it has already been observed that the "functional planning" of O'Malley et al. (1985) has been advocated as a strategy from which SL learners might benefit if taught to use it. This form of planning involves the learner in evaluating what sort of language is needed to complete a given SL task, determining whether he or she has command of that language, and taking steps to learn additional lexical items, plan the use of relevant constructions, and the like. This would seem to be quite a demanding task, though on the face of it, a productive one. The kind of planning used in the present study might also be beneficial to learners and might be capable of being used by less proficient learners or those who are less capable of "functional planning," which requires an analysis of a task which is partially beyond the command of the current IL system.

(Received 2 May 1989)

NOTES

1. I am grateful to my colleague Dennis Preston for drawing my attention to the latter point.
2. Nineteen participants could not report TOEFL scores—these individuals were probably at the lower end of the range on this measure, since those attending University of Hawaii English service courses prior to entry into degree programs do not need to have taken the TOEFL test. These scores are reported here merely to give a general indication of the subjects' English level and were not used as a basis for selection.
3. One of the two productions of each speaker was selected by the toss of a coin. Utterances in this discourse were numbered consecutively, and the discourse was entered using a table of random numbers. At the entry point, five consecutive utterances per speaker were selected for checking.
4. Hubbell (1988) identifies five components in the verb phrase: modal, perfective, continuum, passive, and head. A simple measure of IL range is given by identifying which of these is present in a given sample of speech and standardizing in terms of the number of utterances in the sample. I refer to this measure as VP range/utterance.
5. See Master (1988) and Chaudron and Parker (1988) for summaries of current acquisition sequence work for the article in English.
6. Rosenthal and Rosnow (1984, p. 358) point out that η^2 may be equated with r^2 for tests with one degree of freedom, which was the case with all of these tests. Cohen's (1977) conventional ranges as stated for r then apply.
7. At least one other investigation of SL learning has found cultural differences in the use of a learning strategy, with Asians being less flexible in adopting a vocabulary learning technique than Hispanics (see O'Malley et al., 1985).
8. DiPietro's *strategic interaction* is an approach which utilizes an element of planning, though from a group perspective. It is, however, "in its infancy" (DiPietro, 1987, p. 147)—that is, there is little supporting empirical evidence as yet.

REFERENCES

- Bailey, N., Madden, C., & Krashen, S. D. (1974). Is there a "natural sequence" in second language learning? *Language Learning*, 24, 235–243.
- Beattie, G. W. (1980). The role of language production processes in the organization of behavior in face-to-face interaction. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 69–107). London: Academic.
- Brown, H. D. (1987). *Principles of second language learning and teaching* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Butterworth, B. (Ed.). (1980a). *Language production* (Vol. 1). London: Academic.
- Butterworth, B. (1980b). Some constraints on models of language production. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 423–459). London: Academic.
- Calfee, R. (1981). Cognitive psychology and educational practice. *Review of Research in Education*, 9, 3–73.
- Cancino, H., Rosansky, E., & Schumann, J. (1975). Testing hypotheses about second language acquisition: The copula and negative in three subjects. *Working Papers on Bilingualism*, 6, 80–96.
- Chaudron, C. (1979). *Complexity of teacher speech and vocabulary explanation*. Paper presented at the 13th Annual TESOL Convention, Boston.
- Chaudron, C. (1985). Intake: On models and methods for discovering learners' processing of input. *Studies in Second Language Acquisition*, 7(1), 1–14.
- Chaudron, C., & Parker, K. (1988, March). *Markedness and topic reference in second language development*. Paper presented at the 8th Second Language Research Forum, Honolulu, HI.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic.
- Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement*, 33, 107–113.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic.
- Cohen, P. R., & Ferrault, C. R. (1979). Elements of a plan-based theory of speech acts. *Cognitive Science*, 3, 177–212.
- Crookes, G. (1986). *Task classification: A cross-disciplinary review* (Tech. Rep. No. 4). Honolulu: University of Hawaii, Center for Second Language Classroom Research, Social Science Research Institute.
- Crookes, G. (1988). *Planning, monitoring, and second language development: A review* (Tech. Rep. No. 6). Honolulu: University of Hawaii, Center for Second Language Classroom Research, Social Science Research Institute.
- Danielewicz, J. M. (1984). The interaction between text and context: A study of how adults and children use spoken and written language in four contexts. In A. D. Pellegrini & T. D. Yawkey (Eds.), *The development of oral and written language in social contexts* (pp. 243–260). Norwood, NJ: Ablex.

- Dechert, H. W., Möhle, D., & Raupach, M. (Eds.). (1984). *Second language productions*. Tübingen: Gunter Narr.
- Dechert, H. W., & Raupach, M. (Eds.). (1980). *Towards a cross-linguistic assessment of speech production*. Frankfurt: Lang.
- DiPietro, R. J. (1987). *Strategic interaction: Learning languages through scenarios*. Cambridge: Cambridge University Press.
- Ellis, R. (1985). *Understanding second language acquisition*. Oxford: Oxford University Press.
- Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in Second Language Acquisition* 9, 1-20.
- Faerch, C., & Kasper, G. (1983). Plans and strategies in foreign language communication. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 20-60). London: Longman.
- Foss, D. J., & Hakes, D. T. (1978). *Psycholinguistics*. Englewood Cliffs, NJ: Prentice-Hall.
- Gaies, S. J. (1977). *A comparison of the classroom language of ESL teachers and their speech among peers: An exploratory syntactic analysis*. Unpublished doctoral dissertation, Indiana University, Bloomington.
- Gregg, K. R. (1984). Krashen's monitor and Occam's razor. *Applied Linguistics*, 5(2), 79-100.
- Hakuta, K. (1978). A report on the development of grammatical morphemes in a Japanese girl learning English as a second language. In E. Hatch (Ed.), *Second language acquisition* (pp. 132-147). Rowley, MA: Newbury House.
- Hatch, E. (1983). *Psycholinguistics*. Rowley, MA: Newbury House.
- Hayes-Roth, B., & Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive Science*, 3, 275-310.
- Holmes, V. M. (1984). Sentence planning in a story continuation task. *Language and Speech*, 27(2), 115-134.
- Hubbell, R. D. (1988). *A handbook of English grammar and language sampling*. Englewood Cliffs, NJ: Prentice-Hall.
- Hulstijn, J. H., & Hulstijn, W. (1984). Grammatical errors as a function of processing constraints and explicit knowledge. *Language Learning*, 34, 23-43.
- Hunt, K. W. (1966). Recent measures in syntactic development. *Elementary English*, 43, 732-739.
- Krashen, S. D. (1977). The Monitor Model for adult second language performance. In M. Burt, H. Dulay, & M. Finocchiaro (Eds.), *Viewpoints on English as a second language* (pp. 213-221). New York: Regents.
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.
- Larsen-Freeman, D. (1975). The acquisition of grammatical morphemes by adult ESL students. *TESOL Quarterly*, 9, 409-442.
- Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly*, 12, 439-448.
- Larsen-Freeman, D. (1983). Assessing global second language proficiency. In M. H. Long & H. W. Seliger (Eds.), *Classroom-oriented research in second languages acquisition* (pp. 287-305). Rowley, MA: Newbury House.
- Littlewood, W. (1981). *Communicative language teaching*. Cambridge: Cambridge University Press.
- Long, M. H. (1988). Instructed interlanguage development. In L. Beebe (Ed.), *Issues in second language acquisition: Multiple perspectives* (pp. 115-141). Rowley, MA: Newbury House.
- Master, P. (1988, March). *Acquiring the article system in English*. Paper presented at the 8th Second Language Research Forum [SLRF], Honolulu, HI.
- Meisel, J., Clahsen, H., & Pienemann, M. (1981). On determining developmental stages in natural second language acquisition. *Studies in Second Language Acquisition*, 3(2), 109-135.
- Miller, G. A., Galanter, E. H., & Pribram, H. H. (1960). *Plans and the structure of behavior*. New York: Holt.
- Naiman, N., Frölich, M., Stern, H. H., & Todesco, A. (1978). *The good language learner*. Toronto: OISE.
- O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin*, 97, 316-333.
- Ochs, E. (1979). Planned and unplanned discourse. In T. Givón (Ed.), *Syntax and semantics 12: Discourse and syntax* (pp. 51-80). New York: Academic.
- Ochs Keenan, E., & Bennett, T. L. (Eds.). (1977). *Discourse across time and space* (Southern California Occasional Papers in Linguistics No. 5). Los Angeles: Department of Linguistics, University of Southern California.
- Omaggio, A. C. (1986). *Teaching language in context: Proficiency-oriented instruction*. Boston, MA: Heinle & Heinle.
- O'Malley, J. M., Chamot, A. U., Stewner-Manzanares, G., Kupper, L., & Russo, R. (1985). Learning strategy applications with students of English as a second language. *TESOL Quarterly* 19, 557-584.
- Parrish, B., & Tarone, E. (1986). *Article usage in interlanguage: A study in task-related variability*. Paper presented at the 20th annual TESOL convention, Anaheim, CA.
- Pica, T. (1982). *Second language acquisition in different language contexts*. Unpublished doctoral dissertation, University of Pennsylvania.
- Pica, T. (1983). Methods of morpheme quantification: Their effect on the interpretation of second language data. *Studies in Second Language Acquisition*, 6(3), 69-78.

- Pica, T. (1984). *Procedures for morpheme data analysis in second language acquisition research*. Paper presented at the 18th annual TESOL convention, Houston, TX.
- Pienemann, M. (1984). Psychological constraints on the teachability of languages. *Studies in Second Language Acquisition*, 6(2), 186-214.
- Pienemann, M., & Johnston, M. (1987). Factors influencing the development of second language proficiency. In D. Nunan (Ed.), *Applying second language acquisition* (pp. 45-141). Adelaide: National Curriculum Resource Center.
- Preston, D. R. (1989). *Sociolinguistics and second language acquisition*. New York: Basil Blackwell.
- Rampton, B. (1987). Stylistic variability and not speaking 'normal' English: Some post-Labovian approaches and their implications for the study of interlanguage. In R. Ellis (Ed.), *Second language acquisition in context* (pp. 47-58). Englewood Cliffs, NJ: Prentice-Hall.
- Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral science*. New York: McGraw-Hill.
- Rubin, D. L. (1982). Adapting syntax in writing to varying audiences as a function of age and social cognitive ability. *Journal of Child Language*, 9, 497-510.
- Rubin, J., & Thompson, I. (1982). *How to be a more successful language learner*. Boston, MA: Heinle & Heinle.
- Tannen, D. (1982). Oral and literate strategies in speech and written narratives. *Language*, 58(1), 1-21.
- Tarone, E. (1982). Systematicity and attention in interlanguage. *Language Learning*, 32, 69-82.
- Tarone, E. (1983). On the variability of interlanguage systems. *Applied Linguistics*, 4, 142-163.
- Tarone, E. (1987, October). *Accounting for style-shifting in interlanguage*. Paper presented at the 11th University of Michigan Conference on Applied Linguistics, Ann Arbor, MI.