Language Learning (1974) Vol. 24, No. 2

PREFABRICATED PATTERNS AND THE EMERGENCE OF STRUCTURE IN SECOND LANGUAGE ACQUISITION[1]

Kenji Hakuta[2]
Harvard University

In order to test the hypothesis that the second language learner possesses a series of structurally cohesive learner systems, speech samples of a five-year-old Japanese speaking learner of English were elicited over a 15 month period. Three types of "prefabricated patterns" were analyzed in detail: (1) patterns using the copula, including all allomorphs of *be*; (2) the segment *do you* as employed in questions; (3) the segment *how to* as in embedded *how*-questions. The analysis supported the contention that the subject was operating within a simple learner system involving prefabricated routines.

A number of alternative strategies of approaching the task of acquiring a second language may be available to the learner, and it is conceivable that a learner may employ several strategies at the same time. Evidence will be presented here that suggests a strategy of learning on the surface structure level: learning through rote memorization of segments of speech without knowledge of the internal structure of those speech segments.

The data to be reported here come from a longitudinal study of the untutored acquisition of English as a second language by a five-year-old Japanese girl whom we shall call Uguisu. Her family came to the United States for a period of two years while her father was a visiting scholar at Harvard, and they took residence in North Cambridge in a working-class neighborhood. The children in that neighborhood were her primary source of language input. Uguisu also attended public kindergarten for two hours every day, and later elementary school, but with no tutoring in English syntax. Most of her neighborhood friends were in her same class at school.

She was observed over a period of 60 weeks, from age 5.4, which was five months after exposure to English began, until age 6.5. Every two weeks for at least two hours, spontaneous speech

[1]This is a revised version of a paper presented at the Summer Meeting of the Linguistic Society of America, Amherst, Massachusetts, July, 1974. The research was supported by Grant GS 37931 X from the National Science Foundation to Professor Roger Brown.

[2]I would like to thank Professor Roger Brown and Jill and Peter deVilliers for their encouragement and advice in the preparation of this paper.

was recorded and later transcribed in traditional orthography. It should be noted that, prior to the first sample, in fact from three months after her exposure to English began, I made repeated attempts to gather data, but she produced little speech. It probably would have been possible to elicit speech from her at that time by bombarding her with questions, but I chose to let her begin speaking in a natural environment, which was in a play situation with her peers. Thus, the first sample could be considered her earliest attempts at production.

In the case of a second language learner, we would expect that, with advanced semantic development and yet no form with which to express such thoughts, the need to learn the various linguistic structures of the target language is especially acute. Obviously, in order to be on an equal level with other speakers of the target language community, the learner must induce the "latent structure" of the language and come to grips with the variables contributing to speaking a language with full native proficiency; but until that point, it is conceivable that the learner will employ a strategy which "tunes in" on regular, patterned segments of speech and employs them without knowledge of their underlying structure, but with the knowledge as to which particular situations call for what patterns. They may be thought of as props which temporarily give support until a firmer foundation is built. This argument is based on the consideration that a developed processing span enables memorization of longer speech segments, and segments, for example, like *this is* would be not too different from individual lexical items. Individual lexical items are memorized by rote; for example, Uguisu in Months I thru III used 121 different nouns, 56 different verbs, and 40 different modifiers in 581 utterances, and this is undoubtedly an underestimate of her lexicon. Although estimates vary, the vocabulary in the native language of a 5-year old English speaker is estimated to be about 2,000 (Smith 1926, reported in McCarthy 1954), so the capacity is present.

This sort of phenomenon, of course, is not non-existent in first language acquisition, but the extent to which it seems to occur is slight. Brown (1968, 1973) has coined the term "prefabricated routines" for utterances such as *what's that* or *what dat* which seem to be memorized wholes. He also mentions sentence frames such as *where's* plus a slot into which different noun phrases may be inserted. What we are interested in especially is the latter type of phenomenon, that is, segments of sentences which operate in conjunction with a movable component, such as the

insertion of a noun phrase or a verb phrase. Perhaps they should be called "prefabricated patterns" rather than "routines."

There are three prefabricated patterns which I wish to deal with: (1) patterns using the copula, including all allomorphs of *be*; (2) the segment *do you* as employed in questions; (3) the segment *how to* as in embedded *how*-questions. What all of these forms have in common is that, in the majority of occurrences employed, they appear to be well formed on the surface, but, as will become apparent shortly, there is evidence suggesting that they are not necessarily so.

The task of identifying prefabricated patterns is complex; the data are not uniform and ultimately many prefabricated patterns will escape without positively being identified. For example, forms like *can you*, *where is*, *what's*, *let's*, *isn't it*, and many others can all be thought of as attachments to verb phrases or noun phrases, and there is some evidence for each, but the evidence is too thin.

It should be reiterated at this point that the existence of prefabricated patterns would not imply anything other than the fact that they may be one of many possible strategies employed by the learner. From the very beginning of the development of language, there are many forms of high flexibility, most notably among them the noun phrase and the verb phrase. For purposes of the present study, these, and consequently much of the richness of the data, will have to be set aside.

The first item on our list involves the various forms of the copula *be*, including the negative in the simple form of *not*, without contractions. The different subjects with which they were used are listed in Table 1. But before turning to the numbers in the table, it should be noted that the copula is supplied in practically all instances. Omission occurs only about one percent of the time. This is striking in light of the fact that one of the better-established phenomena in the study of language acquisition is variability while a form is being acquired. For example, grammatical morphemes when scored for presence or absence in obligatory contexts show a high degree of variability during the slow and gradual course of development from total absence of total presence (Brown 1973). Question-inversion also shows similar patterns.

The development of Uguisu's grammatical morphemes has been reported in an earlier paper (Hakuta 1974); the acquisition of these morphemes has been found to be gradual and the performance to be variable. Figure 1 reproduces the acquisition curves for the possessive inflection *'s* and the auxiliary *be* for the future form

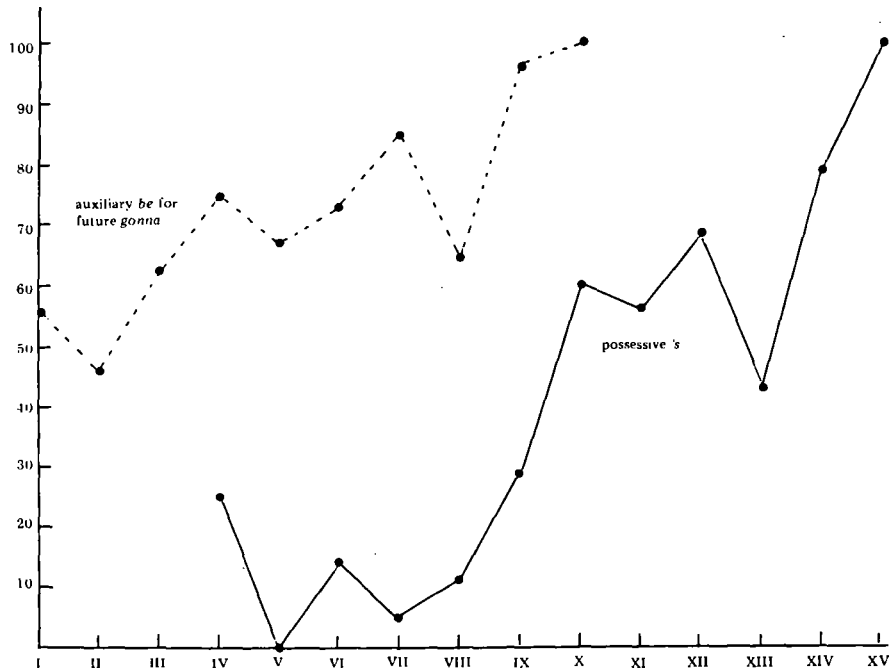

Figure 1. Acquisition curves for two grammatical morphemes in Uguisu.

gonna. Why is such not the case with the copula? If one were to superimpose a line showing the percentage of copulas supplied in obligatory context, it would be a straight horizontal line spanning the length of the graph. This lack of variability makes one question the generative use of the copula.

To return now to Table 1, the numbers in the cells represent the number of instances of occurrence per token of 200 utterances for each one-month period. The 200 token utterances are the combination of the first 100 utterances from two bi-weekly samples. Utterances which were interrupted, or badly recorded as well as one-word replies to questions were omitted from count. The "totals" column indicates that for Month I, out of a token of 200 utterances, 107 of them involved the copula, and in Month II, there were 52. This is the second striking aspect of Uguisu's copula sentences; in the first month, they constituted over half her utterances, and then from the second month on, the percentage drops, eventually to about 20 percent. A sample of 300 utterances from native English speaking peers revealed a steady 20 percent

TABLE 1

Distribution of copula patterns in tokens of 200 whole utterances for each month period

Month	This	That	It	I	You	She	These	Here	NP	We	He	They	Total
I	47	23	11	13	2	1	0	3	7	0	0	0	107
II	17	9	8	7	6	0	1	2	2	0	0	0	52
III	19	4	2	9	4	0	3	2	3	0	0	0	46
IV	13	7	4	0	2	1	0	1	9	1	1	0	40
V	17	2	8	1	0	2	1	0	4	0	0	0	35
VI	8	7	3	8	3	6	1	0	3	0	0	0	39
VII	24	12	7	1	3	3	2	0	2	0	0	0	54
VIII	14	11	12	1	3	4	0	0	5	0	0	0	50
IX	6	6	13	2	0	0	0	0	3	1	0	0	31
X	15	11	8	1	0	2	0	1	1	0	2	1	42

proportion of copula usage throughout the age range. It is interesting to note that Uguisu comes to level off somewhere near that proportion.

The distribution of the copula patterns, as can be seen in Table 1, is quite uneven, mostly centering around *this is*, *that is*, *it's*, *I'm*, and *NP is*. The *NP is* pattern might be considered as different from the rest in the sense that it could be argued that it is being supplied generatively, but it still seems to be patterned since it shows no variability and also, there is lack of number agreement with plural noun phrase subjects. In fact, number agreement did not begin appearing with any reliability until Month X.

The notion of patterns is probably best demonstrated in the case of the pattern *these are*, which began to be used at Month II. Table 2 gives a list of instances of sentences involving *these are*, as well as instances of where *these are* should have been supplied but instead took the form *this is*, along with proportions of correct instances of number agreement with plural noun-phrase subjects and *Wh*-questions with plural subjects.

It is interesting to note that 31 percent of the *these are* patterns used were incorrect, that is, they had singular referents. As can be seen in Table 2, this trend is quite consistent throughout the samples, although more evident between Months III through VII. Furthermore, instances where *these are* should have been used rather than *this is* do not disappear altogether. This brings to light an interesting interplay between form and function. The form in this case is the rigid pattern *these are*, and the function which is to

TABLE 2

Tabulation of singular and plural patterns provided by Uguisu

Month	*This is[1]	These are[2]	*These are[3]	NP Pl. Subj.[4]	Wh-pl. Subj.[5]
I	9	0	0	—	—
II	6	2	0	0/4	—
III	3	24	6	1/9	1/1
IV	1	1	1	0/7 (1 omit)	0/2
V	5	2	5	0/4 (1 omit)	0/4
VI	3	8	7	0/4	0/1
VII	1	7	4	— (1 omit)	—
VIII	4	3	1	—	0/1
IX	1	3	1	1/5	—
X	3	0	0	4/7	0/1
XI	0	11 (1 omit)	5	3/3	—
XII	0	2	0	2/2	—
XIII	0	2	0	4/5	2/5
XIV	0	0	0	—	—
XV	0	3	0	3/3	—

[1]*This is* used when *these are* is obligatory.
[2]*These are* used correctly in proper context.
[3]*These are* used in incorrect context.
[4]Proportions of *are* supplied with Pl. NP Subject.
[5]Proportions of *are* supplied in Wh-questions with plural NP subject.

be expressed is plurality. The general conclusion to be drawn from Table 2 is that when the pattern *these are* is used, it tends to be used in a plural context, but at the same time, it occasionally misses the target and ends up in an utterance such as *these are my house*; or else it is not used and the utterance produced ends up as *this is two cup.*

One final point about *these are*: when the form is permuted, *are* turned into *is.* Thus we have the illustrative contrasting examples,

All these are sick. => All these people is sick.
Why these are dirty? => Why these floor is dirty?

both pairs of which were uttered in the same session, in the latter case within seconds of each other as Uguisu responded to my prompt "What?" Quite interestingly, at month XI, when Uguisu's number agreement was in blossom, in one instance she omitted *are* and in another instance showed a trace of intrusion error:

These is for . . . these are for big person like my, I.

The second pattern to be considered was the question form *do you*. From the very first sample, *do you* appeared in questions such as the following:

Do you know?
How do you do it?
Do you have coffee?
Do you want this one?

This pattern is evidently correct in surface form, and most questions are concerned with the listener, which means that in most cases this pattern matches what should be said. The only problems occurred either when the subject was not *you* or in the past tense usage. At Month II, she produced the following utterances while playing "teacher" and asking friends about a set of pictures:

What do you doing, this boy?
What do you do it, this, froggie?
What do you doing?

Three months later, in a translation task, she said:

What do you drinking, her?

The above four utterances, unfortunately, were the only attempts that Uguisu was to make using a third person subject in this type of question in all of the fifty hours of speech recorded, but she consistently failed. What they suggest is that *do you* has not at all been segmented, but was learned as a question marker.

When does a segment such as *do you* yield its internal structure? Table 3 gives us a glimpse of what may have been involved. It is a list of contexts requiring the past form of *do* when used as an uaxiliary in questions. Initially, it does begin in unmarked form, but one sees a gradual sprinkling of tense-marking going down the list. It should be noted here that, although the evidence is too thin to support a generalization, the development here does not appear to be sudden. At Month IV, the past marking appears with *where*, at Month V with *what*, but not in the case of *how* until Month VI, and even then, there is some variation. With the *do you* in yes-no questions, it occurs even later, at Month VIII, but until that point, it is the only form where past tense marking appears in the main verb. And after Month IX, the past tense marking is reliably present in all cases. This gradual replacement of form is not inconsistent with what is found in the pattern *how to*, to which we now turn.

TABLE 3

Contexts requiring past auxiliary did *in question form*

Month	Unmarked	Marked
III	Why do you do? How do you make? How do you draw that?	
IV	What do you do?	Where did you get that?
V	How do you break it?	What did she say? What did you say? What did you say?
VI	Do you bought too? Do you bought this too? Do you put it? Do you put it? How do you put it? How do you put it?	What did you do? What did you say?
VII	How do you do it?	How did you get it?
VIII	Do you saw these peppermint? Do you saw some star eye? Do you saw some star eye?	Did you call? Did everybody saw some blue hairs?
IX		Did you see the ghost? Did you know we locked the door when we come to here?
X		Did you use some blue? Why did you do that? Why did you get this? Why did you go to a hospital? Why did you draw?
XI		What did you say? What did camel say? Did I made that? Did I make that? Did you see that? Did you see me? Why did you put this? I didn't correct this one, did I?
XII		Did you what?

At Month III, Uguisu made the following set of utterances:

I know how to do it.
I know how to do read it this.
I know how to read it this.
I know how to make.
I know how to draw it cat.
I know how to draw (it) butterfly.
I know how to draw it boy.

What appeared at that time to be a grammatical construction of embedded *how*-questions, however, disintegrated over time into forms such as the following, which she produced at the very last session:

First I gotta write it and show you how do you spell "Debra."
I know how do you spell "Vino."
We only know how do you make it like that.
I know how do you write this.

It is an interesting progression, because we would expect that if she were relying solely on the surface patterns of her peer speech, she would stay with what she had. But she did not; gradually the proportion of correctly used *how to* forms went from nearly 100 percent in months III to VI, to a zero to 50 percent correctness in months XI to XV. Just what is Uguisu doing? We can get a general idea if we look at the development of other *wh*-embedded questions. Table 4 gives a list of all embedded "where" questions from the samples. It begins with the configuration "sentence + question" (see Hatch 1974) with inversion (I don't know where is money) and gradually undergoes metamorphosis where the copula becomes uninverted (I know where it is) after some redundancy (You will see where is your house is). Perhaps what Uguisu did, then, was to initially learn the form *how to* as a unit, (It is possible that the first 7 utterances with *how to* were all acquired in the pattern *I know how to* plus a verb phrase; in later samples, she uses *how to* with other verbs, such as *show*, *tell*, and *be*.) and later replacing it with the form "sentence + question," which would be more consistent with the other forms of *wh*-embedding. Although the study was terminated before Uguisu had the chance to restructure the now deviant form of *how*-embedding, just as she had to restructure her *where*-embedding, it is probable that it would have returned to the proper *how to* form eventually, had she remained in the United States.

TABLE 4

Embedded "where" questions produced by Uguisu

Form*	Month	Sentence
I	I	I don't know, where is money.
I	IV	We know where is this.
I	IV	I don't know where is it.
I	V	My father tell me where is here.
I	V	I didn't know, where is, um, doctor's room.
U	VI	I know where it is.
I	VI	You have to close your eye and you have to see where is it.
U	VI	I don't know where she is.
I	VII	I don't know where is your house.
I	VII	I didn't know where is it.
I	VIII	You know where is my house.
I	VIII	You will see the house where is it.
R	VIII	You will see where is your house is.
R	VIII	I don't know where is the telephone number is.
R	IX	I don't know where is the woods is.
I	IX	I know (it) where is it.
U	IX	I know where it is.
U	XII	I don't know where the bathroom is.
U	XIII	I know where it is.

*I = inverted; U = uninverted; R = redundant.

What has been suggested above by the three items, of course, is extremely limited. However, many interesting and puzzling issues are raised: (1) If Uguisu did not happen to have tuned in on the pattern *how to*, what would her first attempts to produce embedded *how*-questions have looked like? (2) Does the rote memorization of a prefabricated pattern accelerate or decelerate incorporation into the structure? In other words, does the learning of a pattern signal or motivate search for its internal structure, or does it hinder the search because the prefabricated pattern is easy to use? And (3) do prefabricated patterns whose internal structure is finally perceived remain as convenient short-cut routes to production, or are they simply discarded, never to be employed again?

REFERENCES

Brown, Roger. 1968. The development of wh-questions in child speech. Journal of Verbal Learning and Verbal Behavior 7.279-290.

Brown, Roger. 1973. A First Language: The Early Stages. Cambridge, Ma.: Harvard University Press.

Hakuta, Kenji. 1974. A preliminary report on the development of grammatical morphemes in a Japanese child learning English as a second language. Working Papers on Bilingualism 3.18-43. Toronto, Canada: Ontario Institute for Studies in Education.

Hatch, Evelyn. 1974. Second language learning—universals? Working Papers on Bilingualism 3.1-17. Toronto, Canada: Ontario Institute for Studies in Education.

McCarthy, Dorothy. 1954. Language development in children. In L. Carmichael (ed.) Manual of Child Psychology. New York: Wiley & Sons.