Comment on "The insufficiency of a finite state model for verbal reconstructive memory"¹

PHILIP B. GOUGH, INDIANA UNIVERSITY
ERWIN M. SEGAL, ARLINGTON STATE COLLEGE

Abstract

Braine (1965) has contended that, in the acquisition of a certain miniature language, what is learned cannot be represented by a finite state model because it requires the assumption that S invented some of its rules. It is argued that any grammatical description of what S learned requires this assumption, and that the finite state model is inadequate instead because it poorly represents what S does invent.

For some years following the advent of communication theory (Shannon & Weaver, 1949), many linguists and psychologists hoped that language might be described as a finite state (FS) Markov process, a device which generates sentences from left to right, proceeding from element to element according to a given set of transition probabilities. This model had great appeal, for it fit comfortably with ideas, like chaining, familiar to learning theorists, and it permitted the application of some well-developed mathematical techniques to linguistic and psycholinguistic description.

But these hopes were shattered by Chomsky's (1956) demonstration that no FS device could be constructed which would produce all and only the sentences of any language which permitted unlimited self-embedding. Since many natural languages (including English) have this property, it is clear that the linguistic competence of the speaker, his knowledge of his language, cannot be represented by a FS grammar.

But since speakers neither produce nor understand sentences with unlimited self-embedding (in fact, there are severe limitations on the degree of embedding they can tolerate; Miller, 1965), this insufficiency of the FS model, though incontrovertible, is psychologically remote. The FS model would be more immediately discredited if it were shown that it provides an inadequate representation of a speaker's knowledge of some *finite* language. Braine (1965) has claimed to have found just this, that in the acquisition of a certain miniature language, " 'what is learned' cannot be represented by a finite state rules" (p. 291). We think he is right, but not for the right reasons.

Braine presented a set of structured verbal items, amounting to a miniature language L, to Ss whose retention of those items was measured with recall, recognition, and completion tests. L (the items presented to S) may be described by either of the grammars A in Fig. 1. (See Braine, 1965, for explanation.)

These grammars yield exactly the same strings, but they differ in a crucial respect. The first is a FS

grammar (FSG), for it contains only rules of the form $A \rightarrow xB$ (where x is a terminal symbol and B not the initial symbol) and $A \rightarrow x$ (Chomsky, 1963). The second is a *phrase-structure* grammar (PSG), a member of a class of grammars which, since they permit unlimited self-embedding, are more powerful than FSG (though not powerful enough to generate at least one natural language; Postal, 1964).

The important things to note in L are (1) that there are dependencies between a and b , and p and q , which reach across the intervening items, and (2) that the elements $x_7 \dots x_{12}$ (described in PSG-A as class J) occur in the context $a \dots b$, but not in $p \dots q$, while the converse is true of $x_{13} \dots x_{18}$ (class K).

When Braine's Ss were exposed to L, they learned the first of these restrictions; in the recognition test, they seldom accepted strings of the form aXq or pXb . But they did not learn the second; strings described in PSG-A as aKb and pJq were accepted as readily as aJb and pKq .

Evidently the language S learned is not identical to L, and what he learned is not correctly described by the grammars A. It may be described by either of the grammars B. Again, the first is a FSG, the second a PSG. They differ from the corresponding grammars A in precisely the way that what Braine's Ss recognized differed from what they heard; they permit the x items to occur in the contexts $a \dots b$ and $p \dots q$ indiscriminately.

FSG A	PSG A
$S \rightarrow aY, pZ$	$S \rightarrow aXb, pYq$
$Y \rightarrow x_1B, \dots, x_{12}B$	$X \rightarrow I, J$
$Z \rightarrow x_1Q, \dots, x_6Q, x_{13}Q, \dots, x_{18}Q$	$Y \rightarrow I, K$
$B \rightarrow b$	$I \rightarrow x_1, \dots, x_6$
$Q \rightarrow q$	$J \rightarrow x_7, \dots, x_{12}$
	$K \rightarrow x_{13}, \dots, x_{18}$
FSG B	PSG B
$S \rightarrow aY, pZ$	$S \rightarrow aXb, pXq$
$Y \rightarrow x_1B, \dots, x_{18}B$	$X \rightarrow x_1, \dots, x_{18}$
$Z \rightarrow x_1Q, \dots, x_{18}Q$	
$B \rightarrow b$	
$Q \rightarrow q$	

Fig. 1. Phrase-structure (PSG) and finite state (FSG) grammars describing what Braine's Ss heard (A) and learned (B). (Each lower case letter represents a nonsense syllable or paralog.)

It is this difference which lies at the core of Braine's argument of the insufficiency of a FS model. Noting that what his Ss learned *could* be described by FSG-B², he argues that it *should* not be: "Finite state rules are clearly not what is learned in this experiment", for "... a finite state description of what is learned would require the assumption that S invented the additional rules $Y \rightarrow x_{13} B, \dots x_{18} B$ and $Z \rightarrow x_7 Q, \dots x_{17} Q$, which have no exemplars in the exposed strings" (p. 292).

The latter statement is correct. But it does not follow that FSG-B is thereby eliminated as a description of what S learned, or else *all* generative grammars are equally proscribed. For the fact is that *any* rules which S comes to must be "inventions." S's task is like that of the child learning his native language. Like that child, S is not given the *rules* of a language; he encounters a set of utterances, and he must invent, or infer, or induce, a set of rules which those utterances exemplify. Thus the assumption that S must have invented rules is not peculiarly inimical to FSG-B; rather it is requisite to any conception of language learning in grammatical terms.

If all rules are inventions, some are more interesting than others. In the present case, the most interesting are those which generate the items reconstructed, but not heard, by S. In FSG-B, these are the rules $Y \rightarrow x_{13} B$, etc., which Braine disputes. These would be inventions twice over, inventions not only in the sense that they were created, but also in the sense that they had not been created by E. But since the utterances accepted by S differed systematically from what he heard, any accurate description of what he learned must be different from one of what he heard, so that this condition does not uniquely condemn FSG-B. In fact, PSG-B, which Braine *accepts* as a description of what S learned, incorporates the rules $X \rightarrow x_7 \dots x_{18}$; in the terms of PSG-A, these are equivalent to $X \rightarrow J$, K and $Y \rightarrow J$, K , and the rules $X \rightarrow K$ and $Y \rightarrow J$ have no exemplars in the exposed strings.³

The insufficiency of a FS model is not demonstrated by these arguments concerning invention, for they would equally invalidate other grammatical descriptions of what S learned, notably including one which Braine himself judges acceptable. But there are several valid arguments against FSG-B. One concerns how much FSG-B would require S to learn, the others what it says he learned.

S heard 24 different strings; he reconstructed 36. One of the reasons for assuming he learned a grammar is that these strings could not have been mastered by rote in one or two presentations. PSG-B uses 20 rules (involving 26 terms) to generate them, a substantial simplification. But FSG-B requires 40 (involving 83 terms). If S could not learn 24 (or 36) strings, it would seem unreasonable to assume he could master even more rules.

This unparsimonious burden of rules might be justified

if they provided a superior description of what is learned or the course of acquisition. But the FSGs do neither.

First, they provide an inferior description of what S learned. When S learns a language, he learns more than the permissible strings of that language; he also learns their structure. Intuition clearly shows that the strings of L are composed of a frame and a filler, that the constituents of a string axb are $a \dots b$ and x , and operational tests could be devised to translate this intuition into data. (For example, having learned L, Ss might be given new vocabulary entries r and a pair $s-t$, and asked to use them in sentences of L; intuition suggests constructions like arb and sxt .) But FSG-B analyzes axb into a and xb (in the proposed test, it leads one to expect strings like rxb and ast), a counter-intuitive, and thus faulty, structural description.

(It should be noted that PSG-B also fails in this regard, for it fails to associate a with p with q constituents of a construction which does not include the filler; any PSG which excludes discontinuous constituents must fail in the same way. An adequate description would be provided by a transformational grammar which generates frame and filler as successive constituents with PS rules, then permutes the filler into medial position by a transformational rule.)

Second, the FSGs provide an improbable description of acquisition. As Braine notes, if Ss were given sufficient exposure, they would eventually learn L; they would master the restrictions his Ss missed. What they learned would move from grammars B to A, from more rules to less in the FSGs, from fewer to more in the PSGs. S's increasing knowledge of L is surely better represented by the latter.

References

- Braine, M. D. S. The insufficiency of a finite state model for verbal reconstructive memory. *Psychon. Sci.*, 1965, 2, 291-292.
- Chomsky, N. Three models for the description of language. I. R. E. Translation on Information Theory, 1956, IT-2 No. 3, 113-124.
- Chomsky, N. Formal properties of grammars. In R. D. Luce, R. R. Bust, & E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. II. New York: Wiley, 1963. Pp. 323-418.
- Miller, G. A. Computers, communication, and cognition. *Advanc. of Science*, 1965, 21, 417-430.
- Postal, P. Limitations of phrase structure grammars. In J. A. Fodor & J. J. Katz (Eds.), *The structure of language*. Englewood Cliffs: Prentice-Hall, 1964. Pp. 137-152.
- Shannon, C. E., & Weaver, W. *The mathematical theory of communication*. Urbana: University of Illinois Press, 1949.

Notes

1. We are indebted to Martin Braine, Jerry Fodor, Donald Foss, James Greeno, James Jenkins, and Kirk Smith for criticism of the initial version of this comment.
2. Since they learned a finite language, this is necessarily so.
3. It might be objected that since the rules $X \rightarrow x_7 \dots x_{18}$ of PSG-B have exemplifications in L (e. g., ax_7b , $px_{13}q$), whereas the rules $Y \rightarrow x_{13}B, \dots x_{18}B$ and $Z \rightarrow x_7Q, \dots x_{12}Q$ do not, the latter are inventions in a sense in which the former are not. But this begs the question, for $X \rightarrow x_7 \dots x_{18}$ may be thought to be exemplified in L only *after* they are invented (i. e., only after $x_7 \dots x_{12}$ and $x_{13} \dots x_{18}$, which have different privileges of occurrence, have been placed in a single class).

A reply and discussion by M. D. S. Braine will be published in this Journal September 15 (No. 6).