

Some Discriminative Properties of Syntactic Structures

HARLAN LANE AND BRUCE SCHNEIDER

Behavior Analysis Laboratory, University of Michigan, Ann Arbor, Michigan

In a recent article discussing some psychological studies of grammar, Miller (1962) discusses an experiment by Mehler (1963) in which Ss are given a list of sentences, varying both in content and structure, to read and then recall. The errors in recall after the first few trials are rarely due to omissions and semantic confusions; the bulk of the errors occur because S alters the syntactic form of the sentence.

Miller suggests that the process of recall begins with a recoding of the sentence into the kernel plus "some kind of implicit code" of the transformation that is necessary to reconstruct the sentence from the kernel on demand. A maximum of three transformations plus the kernel were needed to reconstruct certain of the sentences that Mehler presented. Thus, Miller proposes that "the sentences were recoded and that each of the four components of the kernel-plus-code was remembered correctly or incorrectly independently of the others." This hypothesis leads to a theoretical distribution of syntactic errors which accords well with that obtained.

A critical step in the recoding of sentences into kernel plus transformation is the discrimination of the syntactic structures involved. One way of exploring the discrimination of syntactic structures is to give a listener some task which requires discriminations among sentences generated by different transformations of the same kernel. In this experimental setting, we may obtain information on the discriminability of syntactic structures, that is, the degree to which they do not overlap in

controlling differential responding. This is the first objective of the present study.

The degree of overlap or "confusion" of syntactic structures may be expected to depend greatly on their relative frequencies within a corpus. If one structure predominates, it may pre-empt discriminative control, and errors in responding will be associated primarily with the less frequent structures. When the several structures approach equal frequency there are three possible outcomes. First, the S may respond appropriately to each. In Mehler's experiment involving recall, Miller reports that "the Ss quickly get the impression that about half the sentences are negative, half are passives, half are questions; in recall, therefore, they try a little probability matching." Second, a particular structure may appropriate control. Third, discriminative control may break down altogether and the S respond randomly with respect to the stimulus array (the corpus).

The second objective of the present study, therefore, is to examine the effects of the relative frequencies of the syntactic structures in a corpus on the discriminative control exerted by those structures.

METHOD

Preparation of the Stimulus Array

Negative, passive, and query transformations were applied to 36 kernel sentences, yielding 144 sentences sampling four syntactic structures. As an example:

The people support Kennedy is a declarative sentence (D).

The people do not support Kennedy is a negative sentence (N).

TABLE 1
 FREQUENCIES OF THE SYNTACTIC STRUCTURES READ BY EACH SPEAKER

Matrix No.	H (Bits)	Speaker No.	Dominant structure	Frequency			
				D	N	P	Q
1	2.0	1	D	36	0	0	0
		2	N	0	36	0	0
		3	P	0	0	36	0
		4	Q	0	0	0	36
		5	D	33	1	1	1
2	2.6	6	N	1	33	1	1
		7	P	1	1	33	1
		8	Q	1	1	1	33
9	3.98	33	D	12	8	8	8
		34	N	8	12	8	8
		35	P	8	8	12	8
		36	Q	8	8	8	12
		37	D	9	9	9	9
10	4.0	38	N	9	9	9	9
		39	P	9	9	9	9
		40	Q	9	9	9	9

Kennedy is supported by the people is a passive sentence (P).

Do the people support Kennedy? is a query (Q).

The 144 sentences were assigned to each of ten 4×4 matrices (Table 1). The rows of each matrix represent 4 speakers, the columns represent the 4 syntactic structures, and the cells give the number of sentences of each form uttered by a single speaker. Each speaker read every one of the 36 kernel sentences in one of its 4 forms. In matrix 1, speaker 1 read all D-sentences, speaker 2 all N-sentences, etc. In matrix 2, speaker 5 read 33 D-sentences and one each of the other forms; similarly, speakers 6-8 read the corresponding dominant structure (N, P, Q, respectively) 33 times and the nondominant structures once each. In successive matrices in the series of 10, the relative frequency of the dominant structure in each row decreased in steps of three, while the nondominant structures increased in frequency in steps of one. The uncertainty of the stimulus array ranged from 2 to 4 bits.

The 36 sentences comprising the corpus of each speaker were arranged in random order both with respect to the distribution of kernels and the distribution of syntactic structures determined by the matrix. Forty male speakers (4 speakers for each of the 10 matrices) read the sentences with conversational intonation while seated in front of a microphone in a radio studio. Each corpus was recorded

on a loop of magnetic tape, labeled, and cartridge separately.

Subjects and Procedure

Six groups of 5 Ss each served in separate sessions lasting 5 hours. A 15-min. hiatus was provided after 2 and 4 hours; Ss were not permitted to discuss the experiment.

The four corpora comprising a single matrix and the 10 matrices were presented in random order to the groups of Ss.

After listening to a set of 4 speakers from one of the matrices, S was given a deck of index cards containing 144 different sentences, representing 36 occurrences of each syntactic structure in random order. Ten such decks were prepared from 1440 simple sentences; a new deck was selected at random and presented to S after each matrix. The Ss were instructed to sort the deck of sentences into 4 piles, each pile corresponding to one of the 4 speakers they had just heard. They were told to assign each sentence to the speaker "you think was more likely than the other speakers to have said it." The Ss were visually isolated from each other and were not constrained in the time to complete the sorting task. The cards were collected so as to preserve the assignment of sentences to speakers.

Ten response (R) matrices were prepared later for each S. The matrices showed the frequency with

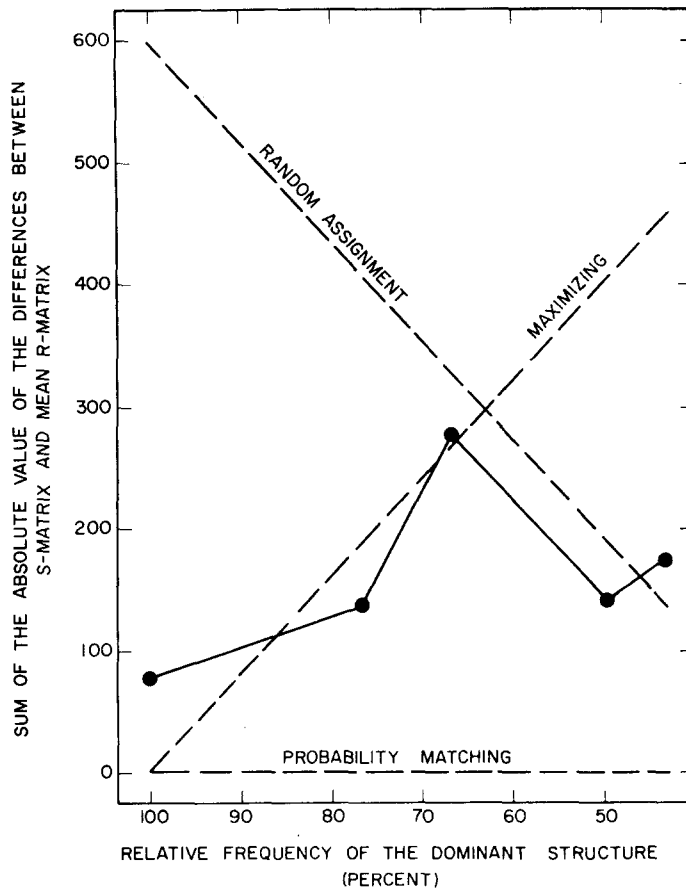


FIG. 1. The effect of the relative frequencies of four syntactic structures in a speaker's corpus on the accuracy of discriminative responding by a listener. The difference between the distributions of structures in the stimulus matrix and in the corresponding average response matrix is plotted as a function of the relative frequency of the dominant syntactic structure in the stimulus matrix. The dotted lines indicate hypothetical results (see text).

which each of the 4 sentence structures was assigned to each of the 4 speakers in the corresponding stimulus (S) matrix. Comparison of the S- and R-matrices revealed the degree of discriminative control, while comparison of the cells within the R-matrix revealed the direction and extent of confusions among syntactic structures.

RESULTS AND DISCUSSION

Degree of Stimulus Control

Figure 1 shows the effect of the relative frequencies of the syntactic structures in the speakers' corpora on the discriminative control of responding by the listeners. The dependent variable is the degree of the mismatch between

the S- and R-matrices for each group of 5 Ss, expressed as the sum of the absolute value of the differences between the matrices.

If the listener's assignment of sentences to speakers was entirely under discriminative control, the distribution of structures in the R-matrix would be the same as that in the S-matrix and the sum of their absolute differences would be zero over all S-matrices. This hypothetical outcome is shown in Fig. 1 by the line labeled "probability matching." A second possible finding, labeled "random assignment," would be obtained if the listeners' assignments were not under discriminative

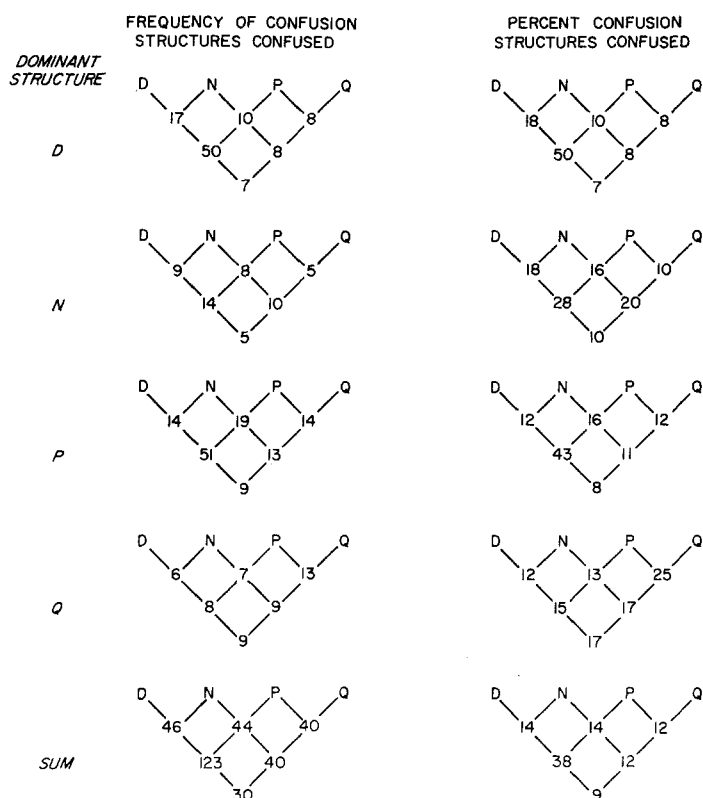


FIG. 2. Patterns of confusion in the discrimination of syntactic structure. The frequency of a confusion between two or more structures is the number of occasions on which 10 or more sentences of each structure were assigned to the same speaker. The dominant structure is the syntactic form uttered most frequently by the speaker when the confusion occurred.

control at all. The third hypothetical outcome, labeled "maximizing," would be obtained if the listeners assigned all 36 sentences with a given structure to the speaker who said that structure most frequently.

The obtained error function, which compares the *first* S-matrix presented to each group with the mean of the corresponding 5 R-matrices, is nonmonotonic and peaks at stimulus matrices of intermediate uncertainty. Comparison of the obtained and hypothetical functions suggests that listeners assigned all occurrences of a given structure to the speaker for whom that structure was dominant, when the dominance in frequency for one structure was marked. However, when the relative frequencies of the four syntactic structures in the S-matrix were more nearly equal, discriminative control was minimal and the struc-

tures in the R-matrix were distributed essentially at random. Examination of the other 9 R-matrices for each listener supports this generalization, although order effects somewhat confound the analysis.

It is interesting to note that the transition from maximizing to random assignment occurred near the point of intersection of the two hypothetical error functions. Had the dominant structure continued to pre-empt discriminative control, a greater mismatch between the S- and R-matrices than that produced by random assignment would have resulted.

Overlap in Stimulus Control

Two or more syntactic forms in the corpus of one speaker were classified as "confused" if the listener assigned 10 or more sentences of

each form to that speaker. For example, if Speaker 15 read 27 P-sentences and 3 each of D-, N-, and Q-sentences, and a listener made the assignments: 20P, 15D, 4N, and 4Q, then P and D are said to be confused when P was the dominant structure.

Figure 2 shows the pattern of confusions when D, N, P, and Q were the dominant structures. When D was dominant, 50% of the confusions that occurred were associated with D and P. Similarly, when P was dominant, most of the confusions occurred with D and P. When N was dominant, the greatest number of confusions still occurred between D and P, although when Q was dominant, Q generalized with P most often.

Table 2, which summarizes these findings, shows the per cent of confusions associated with each syntactic structure as a function of

TABLE 2
THE SYNTACTIC STRUCTURES WHICH WERE CONFUSED
WHEN A DOMINANT STRUCTURE WAS SENT

Structures confused	Per cent confusion				Total
	Dominant structure				
	D	N	P	Q	
D	12	4	12	4	32
N	5	4	7	3	19
P	11	4	13	4	32
Q	4	3	6	5	18
Total	32	15	38	16	

the dominant structure in the speaker's corpus. As might be expected, most of the confusions that occurred in assigning sentences to a speaker involved the syntactic structure read most often by the speaker. It is interesting to note, however, that P-sentences were confused with all other structures as often as D-sentences when D was dominant, and, likewise, there were almost as many D-confusions as there were P-confusions when P was dominant.

The column marginals of the per cent-confusion table show that nearly twice as many confusions occurred when D or P was dominant in a speaker's corpus than when N or Q was dominant. Collapsing over the dominant

structures read by the speakers, the row marginals show that D- or P-sentences were more often confused with the other three structures, that is, more often assigned along with them to the same speaker, than were the N- or Q-sentences. Q-sentences were least often confused with the other structures.

SUMMARY

Forty scripts were prepared that contained the same 36 kernel sentences with one of four syntactic structures: declarative, negative, passive, or query. The relative frequency of the structures in each script was varied systematically. Tape recordings of the scripts, read by 40 speakers, were presented to listeners who sorted a (different) set of 144 sentences according to their estimate of which speaker was most likely to have said each sentence. The distribution of syntactic structures in the stimulus scripts was compared with the corresponding distribution arrived at by the listener.

When a particular syntactic structure predominated in a speaker's corpus, the listeners tended to assign all the sentences of the dominant form exclusively to that speaker. When the relative frequencies of the syntactic structures in a corpus were more nearly equal, the listeners tended to assign structures at random. Most of the confusions in discrimination, that is, assignment of a structure to a speaker much more often than it was uttered, were associated with the declarative and passive sentence structures. These syntactic forms generalized most often with each other and more often with the remaining syntactic forms than either negative or interrogative sentences.

REFERENCES

- MILLER, G. A. Some psychological studies of grammar. *Amer. Psychol.*, 1962, **17**, 748-762.
MEHLER, J. Some effects of grammatical transformation on the recall of English sentences. *J. verb. Learn. verb. Behav.*, 1963, **2**, 346-351.

(Received May 15, 1963)