The comprehension and verification of ambiguous sentences[1]

DONALD J. FOSS,[2] *UNIVERSITY OF TEXAS AT AUSTIN*, THOMAS G. BEVER, *ROCKEFELLER UNIVERSITY*, AND MAURY SILVER, *HARVARD UNIVERSITY*

When Ss are presented with an ambiguous sentence they tend to interpret it in only one way. If later events warrant, Ss can recover the other meaning, a process which takes time. These conclusions follow from the results of a study in which 40 undergraduate Ss verified whether or not pictures shown at the end of a sentence represented the meaning of the sentence. When ambiguous sentences were presented, the verification time (VT) was no slower than for unambiguous sentences if the picture represented the "expected" meaning (as determined on a pre-test) of the ambiguity. The VT to the picture representing the "unexpected" meaning of the ambiguity was longer than VT to corresponding control sentences.

At times listeners notice that an ambiguous sentence is ambiguous, i.e., they interpret the sentence in more than one way and can state this fact. However, in normal discourse people do not usually perceive more than one meaning of the sentences they hear—even though many everyday sentences are in fact ambiguous. These two observations are relevant for a model of sentence comprehension since they pose a problem about the number of grammatical analyses that are oridinarily imposed on sentences.

There are a number of alternative models of this process (MacKay, 1966). (1) Listeners fully analyze all the possible syntactic structures and semantic interpretations and then choose among them by some procedure (e.g., utilizing the preceding or later sentence context). (2) Listeners hold the unanalyzed string in abeyance, not assigning any structure or meaning to it until further information permits a single interpretation. (3) Listeners immediately compute only one structure and meaning and maintain it unless further input necessitates a change or recomputation of the sentence structure and meaning. While the last of these models seems to accord best with intuition, some data have been cited to support Model (2) (MacKay, 1966).

MacKay (1966) found that completion times for sentence fragments were longer when the fragment contained an ambiguity than when it did not. This finding is consistent with either Models (1) or (2). If (1) were true then the increased completion time for ambiguities might be due to competition between the two meanings; if (2) were true then the increased completion time for ambiguous sentence fragments would be due to the lack of any confirmed interpretation. MacKay opted for the latter interpretation because he found that the types of ambiguities which yielded the longest completion times were also the types that were the hardest to discover in an ambiguity detection experiment (MacKay & Bever, 1967). If the completion time effect were due to competition between two meanings, MacKay argued, then the sentences in which the two meanings are most obvious ought to have the largest effect on completion times rather than the smallest effect.

There are, however, some reasons for submitting the question of immediate perceptual processing of ambiguities to another test. Since the completion times in the MacKay (1966) study were on the average more than 7.2 sec, there is some question whether that study was tapping normal mechanisms of immediate sentence comprehension. For example, the relatively high completion time for ambiguous sentence fragments may be due primarily to the long times involved in all sentence completion. That is, given that the S takes a long time to process the stimulus sequence anyway, he may eventually develop both meanings of an ambiguous sequence and then have to decide among them. Furthermore, in their ambiguity detection experiment MacKay and Bever found that Ss easily identified which meaning of an ambiguous sentence they had interpreted first. If Model (2) were correct, this sort of decision should be extremely difficult. In addition, the general view that Ss respond to an ambiguity by assigning it no interpretation involves an apparent contradiction, since by definition Ss would have to know that two interpretations are possible to know that an ambiguity existed. For these reasons the results of the previous studies do not unequivocally support any of the models for normal immediate sentence comprehension.

We assume that when S is presented with an ambiguous sentence with no disambiguating context he typically selects one of the possible structures, i.e., we assume that Model (3) is correct for normal sentence comprehension, which usually occurs within a second. This perceptual strategy may lead to difficulty for S later on, however, since he might later be presented with a situation in which the other structure or meaning must be considered in order to make an appropriate judgment. On the other hand, if S has not made any immediate interpretation (Model 2), or if he maintains both interpretations in memory (Model 1), then he will not have any differential difficulty in making a judgment immediately subsequent to the sentence.

In the present experiment S was asked to verify whether a picture that was presented at the end of an ambiguous or non-ambiguous sentence was a pictorial instantiation of the meaning of the sentence, i.e., he was asked if the picture was "right" or "wrong" given the meaning of the immediately preceeding sentence. Corresponding to each ambiguous sentence there were two pictures. If S interpreted the ambiguous sentence in one way but then saw a picture depicting the other meaning, his response time to verify that the picture was an instance of one of the meanings of the sentence ought to be relatively long since he would have to reinterpret the sentence, a process which presumably takes time. On the other hand, if S was holding both, or neither, of the interpretations, then no *re*interpretation would be required and the time to respond "right" to an ambiguous sentence ought not to be affected by which picture was presented.

METHOD

Materials and Design

Seventy-two basic sentences were constructed. Sixty of these were non-ambiguous sentences and were the same across all Ss. These sentences were all 7 to 10 syllables in length and in the active voice. Each described some readily picturable situation or event (e.g., "The rocket is on the launch pad.").

There were 12 ambiguous sentences, equally divided among the three basic ambiguity types, lexical, surface structure, and deep structure ambiguities. A lexical ambiguity occurs when a single word can have more than one meaning within the sentence while the grammatical structure remains unchanged (e.g., the word "pipe" in "The man is holding a pipe."). A surface structure ambiguity occurs when one string of words can be grouped together in two meaningful ways [e.g., "The boy is looking up the street," can be either (The boy) (is looking) (up the street) or (The boy) (is) (looking up) (the street)]. Such groups are usually indicated by intonation which can, therefore, disambiguate a surface structure ambiguity. Intonation cannot disambiguate the other two types. A deep structure ambiguity occurs when, roughly, the words and phrases of the sentence can enter into more than one set of grammatical relations. The grouping of the words can remain the same [e.g., in "The elephant is ready to lift," the phrase "the elephant" can either be the logical subject (soon the elephant will have lifted something) or the logical object (someone has just readied the elephant to lift it someplace)]. See MacKay and Bever (1967) for further discussion of these types.

Each of the 12 ambiguous sentences had two unambiguous sentences paired with it. These differed minimally in structure and meaning from the ambiguous versions. Each of these two

Table 1
Example of the Stimulus Materials

	Picture 1 (Boy and Road)	Picture 2 (Boy and Map)
Ambiguous Sentence	The boy is looking up the street.	The boy is looking up the street.
Unambiguous Sentence	The boy is gazing up the street.	The boy is looking up the address.

unambiguous control sentences was correctly pictured by one of the two pictures which represented the meanings of the ambiguous sentence. Thus, each interpretation of the ambiguous sentence had an unambiguous counterpart which received the same picture as it did. (See Table 1 for a paradigm of the stimulus materials.)

The pictures were photographic slides of ink sketches of the activities outlined in the sentences. All of the ambiguous sentences were followed by a picture that was right on *one* of the interpretations of the ambiguity. The pictures for slightly more than half of the unambiguous sentences were wrong in order to balance the total number of right and wrong pictures in the experiment. These pictures differed from the sentences with respect to subject, verb, or object. In most cases the incorrect picture differed from the sentence in only one of these respects.

The ambiguous sentences were chosen from a set that had been pretested to obtain a measure of "meaning bias." Meaning bias is a percentage measure of the degree to which one interpretation of the ambiguous sentence is thought of first or expected by Ss. In the pretest 20 Ss heard a list of 24 ambiguous and 24 unambiguous sentences (in a flat tone of voice) and were asked (1) if each sentence was ambiguous (examples were given to insure that all Ss had the appropriate concept), and (2) if so, which interpretation of the ambiguity had occurred to them first. For the experiment proper, 12 sentences were assigned an expected and an unexpected meaning on the basis of this pretest.

Since only one version of each ambiguous sentence could be shown to each S, the sentences were divided into two groups, six were presented in the expected version and six in the unexpected version to one group of Ss; the pairings were reversed for the second group.

Differences in response times to expected and unexpected versions of an ambiguous sentence cannot be directly compared, however, since such differences might be due in part to the time needed to visually search the two different pictures that were presented. Therefore, two more groups of Ss were studied. The first received the same set of pictures as the first experimental group above, but were presented the appropriate unambiguous counterpart of each ambiguous sentence. The second was matched with the second experimental group in just the same way.

If Model (3) is correct, then the difference between verification time (VT) for an unexpected picture following an ambiguous sentence and the VT for the same picture following the corresponding unambiguous sentence ought to be greater than the difference between VT for the expected picture following that ambiguous sentence and the VT for that same picture following its corresponding unambiguous sentence.

Subjects

The Ss were 40 undergraduates (10/group) at Harvard and Radcliffe Universities. They were paid for their participation in the 1-h experiment. The 20 Ss in the pretest were obtained from the same source.

Procedure

Each S was instructed that he was participating in an experiment in the understanding of normal English sentences. He was told to judge whether a picture that appeared at the end of each sentence did or did not depict what the sentence said. He was instructed to make his decision and to say "right" or "wrong" as quickly and as accurately as he could.

An example was shown and then the 72 sentences were presented in sequence from a tape in a flat tone of voice. The pictures were shown on a rear projection screen so that E was not visible to S throughout the study. The time between the beginning of one sentence and the next was about 25 sec.

At the end of each sentence the picture was projected on the screen and simultaneously a timer was started. The timer was stopped by a voice-operated relay which operated when S said "right" or "wrong." The first ambiguous sentence occurred on Trial 22, to permit most of the warmup effects to occur. Thereafter, at least three unambiguous sentences intervened between any two ambiguous sentences. At the end of the experiment S was questioned about which version of the ambiguous sentences he had expected, if any.

RESULTS

The latency measures were only scored if S gave the correct response to the picture. Overall, errors were made on 1.3% of the unambiguous sentences and 19.2% of the ambiguous sentences. Significantly more errors were made on ambiguous sentences when followed by unexpected pictures (27.5%) than when followed by expected pictures (10.8%), $t(19) = 4.35$, $p < .01$.

In order to reduce skewness in the latency data and more readily to compare Ss, the data from each S were normalized by a log transform. Further, the means of each of the four groups were set close to zero (to remove any overall chance differences in VT between groups) by subtracting the mean log latency of the unambiguous sentences in each group from all sentences in that group. All further analyses were performed on these transformed data. See Table 2 for the relevant means.

Sentence by sentence all groups performed quite comparably on VT for unambiguous sentences. For example, the product-moment correlation for VT on unambiguous sentences was .78 between one experimental group and its control and .80 between the other experimental group and its control. Combining the two experimental groups and the two control groups yielded a correlation of .88.

Two distributions of difference scores were obtained by subtracting the mean log latency of each sentence in the control groups from the mean log latency of each corresponding sentence (including ambiguous ones) in the appropriate experimental groups. The differences between these sets of scores were then computed, sentence by sentence. The mean of this latter distribution was .024, the variance was .016. The latter figure was used as the estimator of the variance of the distribution in all further tests.

When only the ambiguous sentences and their controls were used in this computation, unexpected ambiguities took longer to respond "right" to than did expected ambiguities. The mean was .091, $t(11) = 1.85$, $p < .05$ (one-tail). This reflects a difference of about .10 sec in the latencies, approximately 1.20 sec for expected and 1.30 sec for unexpected ambiguities (these figures were obtained from the original data).

The measure of bias for one of the surface structure ambiguous sentences was grossly in error according to the posttest questioning. This was not surprising since these ambiguous sentences are very sensitive to slight intonational changes and recording the sentences in a "flat" voice does not eliminate all such cues. Apparently the intonation for this sentence was slightly changed from the pretest since 18 of the 20 Ss who heard the sentence reported that they had expected the version which was classified unexpected on the basis of the pretest. If we change expected to unexpected and vice versa for this one sentence the

Table 2
Transformed Means of VT by Sentence and Picture Types

	Expected		Unexpected	
	Ambiguous	Unambiguous	Ambiguous	Unambiguous
Lexical	.139	.220	.296	.181
Surface	.384	.351	.332	.275
Deep	.268	.262	.310	.252

above mean becomes .130, $t(11) = 2.93$, $p < .02$. We extended this latter analysis and assigned each ambiguous sentence for each S to the expected or unexpected category on the basis of his posttest responses. When this is done the difference between unexpected and expected ambiguities is .177, $t(11) = 4.23$, $p < .01$.

Although the numbers of sentences are small, it is possible to compare the results by type of ambiguity. Most of the overall difference between unexpected and expected pictures was due to the lexical ambiguities, $\bar{X} = .197$, $t(3) = 2.76$, $p < .05$. Neither surface structure nor deep structure ambiguities reached significance although both types were in the same direction.[3]

Considered alone, the VT for ambiguous sentences with expected pictures was slightly shorter than the VT for their unambiguous control sentences, $\bar{X} = -.01$, $t(11) = 1.05$, $p > .10$. The VT for ambiguous sentences with unexpected pictures was slightly longer than the VT for their unambiguous controls, $\bar{X} = .08$, $t(11) = 1.46$, $p < .10$ (one-tail). The present data also permitted a determination of whether it takes longer to verify an incorrectly pictured unambiguous sentence than a correctly pictured one. Gough (1965) and Slobin (1966) have reported that VT is longer for incorrectly pictured sentences. In this study no difference between "rights" ($\bar{X} = -.062$) and "wrongs" ($\bar{X} = -.064$) was found.

DISCUSSION

The results support a model of normal sentence comprehension which states that Ss typically assign only one immediate interpretation to an ambiguous sentence. Only if that interpretation is found to be incorrect does S reinterpret the sentence. We do not have a satisfactory model for making specific predictions about interpretation bias, i.e., about which of the meanings of a sentence will be immediately computed by a particular S. However, the main effect is increased when we take into account which meaning of an ambiguous sentence S himself expected, rather than assigning the "expected" meaning on the basis of the pretest.

The VT differences between unexpected and expected pictures relative to their controls was longer for lexical ambiguities than for surface and deep structure ambiguities. In the MacKay and Bever (1967) study, Ss detected the presence of lexical ambiguities most quickly. It might have been expected, then, that the results of the present study would have been the reverse of what they were, i.e., that Ss would recompute the meanings of lexically ambiguous sentences most quickly when presented with an unexpected picture. However, as MacKay and Bever recognized, the process of searching for an ambiguity may be quite arbitrary with respect to normal comprehension strategies. (For example, finding a lexical ambiguity when S knows there is an ambiguity may be relatively easy because of its specificity.) Also, the effect observed here must be considered tentative because of the small sample of ambiguous sentences used in the study. In addition, it might be the case that the difference between the two pictures is in some sense greater for lexical ambiguities than for other types (see below). If so, the present difference in the effects due to different types of ambiguities would be an artifact of the verification procedure.

When Ss were faced with an unexpected picture following an ambiguous sentence they tended to say "wrong" more often than when an expected picture was presented. This in itself is evidence for Model (3). We must ask, however, why Ss do not always say "wrong" when presented an unexpected picture. The present results could be accounted for by postulating that Ss made an implicit high-speed description of the picture. When the description matched the analysis of the sentence, S responded positively, otherwise negatively. In the case where the description of the picture matched the surface of the sentence analysis (e.g., the lexical items were represented in the picture) but did not match the full analysis, S might attempt to reconstruct the sentence. This reinterpretation would take time and thus account for the basic data. It would also account for the relatively high effect of lexical ambiguities on VT, since it is in these cases that even the expected lexical analyses are themselves changed in the unexpected interpretation.

The main results of this study are summarized as follows: (1) Errors in verification occurred significantly more often when an unexpected picture was presented after an ambiguous sentence than when an expected picture was presented. (2) There was a difference in VT for expected and unexpected alternatives of an ambiguous sentence. The VT for the latter minus its control was longer than VT for the former minus its control. (3) The expected meaning of an ambiguous sentence had no slower VT than an unambiguous sentence. Thus, ambiguity per se does not seem to interfere with understanding the meaning of a sentence. These results support a model of normal sentence comprehension according to which the S at first computes only one meaning of an ambiguous sentence.

REFERENCES

GOUGH, P. B. Grammatical transformations and speed of understanding. Journal of Verbal Learning & Verbal Behavior, 1965, 4, 107-111.

MacKAY, D. G. To end ambiguous sentences. Perception & Psychophysics, 1966, 1, 426-436.

MacKAY, D. G., & BEVER, T. G. In search of ambiguity. Perception & Psychophysics, 1967, 2, 193-200.

SLOBIN, D. I. Grammatical transformations and sentence comprehension in childhood and adulthood. Journal of Verbal Learning & Verbal Behavior, 1966, 5, 219-227.

NOTES

1. This study was supported by NIGMS Training Grant No. 5T01-GM-01011-05 to Harvard University, Center for Cognitive Studies, and by NSF-USDP Grant GU-1598 to the University of Texas at Austin. The authors would like to thank Harris Hartz for assistance in gathering and analyzing the data.

2. Address: Department of Psychology, Mezes Hall, University of Texas, Austin, Texas 78712.

3. One-half of the Ss in each group received all the sentences in the left ear, one-half in the right ear. The data reported in the body of the text are summed across both ears. An analysis of the results by ear (but with the overall group means subtracted out, rather than the group means by ear) indicated, albeit with small Ns, that the differential VT results occurred primarily when the sentences were heard in the right ear ($\bar{X} = .219$). The effects for the left ear ($\bar{X} = .029$) were tenuous. By ambiguity type, lexical, surface structure, and deep structure, the means (computed as stated in the body of the text) for the right ear were .257, .299, and .101, respectively; while those for the left ear were .131, .018, and −.063, respectively.

(Accepted for publication August 9, 1968.)