COGNITIVE SCIENCE 3, 213–230 (1979)

An Evaluation of Story Grammars*

JOHN B. BLACK
Yale University
AND
ROBERT WILENSKY
University of California, Berkeley

We evaluate the "story grammar" approach to story understanding from three perspectives. We first examine the formal properties of the grammars and find only one to be formally adequate. We next evaluate the grammars empirically by asking whether they generate all simple stories and whether they generate only stories. We find many stories that they do not generate and one major class of nonstory that they do generate. We also evaluate the grammars' potential as comprehension models and find that they would add nothing to semantic models that focus on the story content. Hence we advocate a story content oriented approach to studying story understanding instead of the structural story grammar approach.

The study of how to understand stories has become a very active area of research in psychology and artificial intelligence. The largest group of story understanding theories in psychology are the "story grammar" theories. In this paper, we argue that, while the "story grammar" theories have contributed some insights, their orientation is unlikely to yield further progress. We propose that a more productive research orientation is exploring the kinds of knowledge needed to understand story content.

Lakoff (1972) proposed the first story grammar. He reformulated Propp's (1968) theory of Russian folktales using the predominant formalism of modern linguistics—namely, using rewrite rules. Colby (1973) also used rewrite rules in

*This research was supported by research grants MH-13905 from the National Institute of Mental Health to Gordon Bower and contract N00014-75-C-1111 from the Office of Naval Research to the Yale Artificial Intelligence Project. Requests for reprints should be sent to either Dr. John B. Black, Yale University, Department of Psychology, Box 11A Yale Station, New Haven, CT 06520 or Dr. Robert Wilensky, Department of Electrical Engineering and Computer Science, Computer Science Division, University of California, Berkeley, CA 94720.

designing a specialized grammar of Eskimo folktales. However, Rumelhart (1975) proposed the first general grammar that was designed to apply to more than a small restricted set of stories. Other general story grammars have been proposed by van Dijk (1975), Thorndyke (1977), Mandler and Johnson (1977), Stein and Glenn (1979), and Johnson and Mandler (in press). Our interest is story understanding in general, so we will evaluate these general story grammars, not the specialized ones.

We use Thorndyke's grammar as an illustration of these theories because it is the simplest. This grammar's first rewrite rule is

$$\text{Story} \rightarrow \text{Setting} + \text{Theme} + \text{Plot} + \text{Resolution}$$

This rule states that a story is composed of a Setting, a Theme, a Plot, and a Resolution. The next rule further specifies the Setting

$$\text{Setting} \rightarrow \text{Characters} + \text{Location} + \text{Time}$$

Thus, a Setting consists of Character, Location, and Time descriptions. The Theme is then defined by

$$\text{Theme} \rightarrow (\text{Event})^* + \text{Goal}$$

That is, the Theme is a Goal State with, optionally, one or more Events that initiate it. The next rule is

$$\text{Plot} \rightarrow \text{Episode}^*$$

which states that a Plot is a series of Episodes. An Episode is then

$$\text{Episode} \rightarrow \text{Subgoal} + \text{Attempt} + \text{Outcome}$$

and an Attempt is

$$\text{Attempt} \rightarrow \text{Event}^* \text{ or Episode}$$

Thus, the Plot is a series of Episodes, each of which is a Subgoal State, one or more Attempts, and an Outcome. An Attempt is a series of one or more Events or an entire Episode. Further rules state that Outcomes and Resolutions can be Events or States but that Goals, Subgoals, and Character, Location, and Time descriptions must be States.

Although Thorndyke's theory is typical, the various story grammars differ somewhat. The major difference between the other grammars and the Thorndyke one is that the others have semantic relations in addition to syntactic relations like those described above. Thus, for example, the syntactic part of the Rumelhart grammar states that a story has a Setting and an Episode, while the semantic part states that the Setting "ALLOWS" the Episode—i.e., the Setting establishes the conditions needed for the Episode to occur. For another example, Mandler and Johnson state that there are Attempts and Outcomes, but go beyond Thorndyke by demanding that the Attempts "CAUSE" the Outcomes. These semantic relations are not necessary for our initial evaluations of the story grammars, but we

will return to them in a later section of this paper when they become relevant.

The story grammar theories make three general claims: the first is that they are grammars that distinguish stories from nonstories, the second is that they are models of story comprehension, and the third is that they are models of memory for stories. Black (1979) and Black and Bower (in press) have evaluated them as memory models and found that they did not predict which story statements readers would remember. Therefore the story grammars are not useful memory models, so we will evaluate them as grammars and comprehension models. Our "grammar" analysis has two parts: the first part is a formal analysis that examines the form of the grammar rules and evaluates them using results from mathematical linguistics; the second part is an empirical analysis that examines various stories and nonstories asking whether the grammars distinguish them.

FORMAL EVALUATION AS GRAMMARS

In this section we look only at the form of the story grammars: that is, we examine the mathematical implications of the theories as particular kinds of abstract symbol manipulation systems. The implications come from applying results from the theory of formal languages (see Hopcroft & Ullman, 1969; Wall, 1972; and Bach, 1974).

All the story grammars are defined by rewrite rules. In general, a rewrite rule has the form

$$A \rightarrow B$$

where A and B can represent anything. This rule states that B can replace A whenever A occurs or, in other words, A can be rewritten as B. The particular theories under consideration here, however, do not have rewrite rules that are this general. In fact, all except one (noted below) are of the form

$$A \rightarrow B1\ B2 \ldots Bn$$

where A is a single nonterminal symbol (i.e., a symbol that can be further rewritten), and the Bi are any number of terminal symbols (i.e., symbols that cannot be further rewritten) or nonterminal symbols. Note that this type of rule rewrites only one symbol at a time (namely A) and that the rewriting is independent of the context in which the symbol occurs (that is, A appears by itself on the left side of the rewrite rule).

Some terminology from formal languages will help in distinguishing the various kinds of rewrite rules. A phrase structure grammar is a set of rewrite rules each of which rewrites only one symbol at a time. There are two kinds of phrase structure grammars, context-free grammars (CFGs) and context-sensitive grammars (CSGs). In a CFG, only one symbol appears on the left side of the

rule, as in the example above. In a CSG, on the other hand, more than one symbol appears on the left side, but still only one is rewritten at a time. For example,

$$A\ B\ C\ D\ E \rightarrow A\ B\ F\ D\ E$$

rewrites C as F in the context of being preceeded by A B and followed by D E. Note that in both of these types of phrase structure grammars the number of symbols on the left side is less than or equal to the number on the right side. This condition must hold for a rule to be a phrase structure rule. Thus, for example, a deletion rule like

$$A\ B\ C\ D\ E \rightarrow A\ B\ D\ E$$

is not a phrase structure rule. The regular or finite state grammar (FSG) is another important kind of grammar. In a FSG each rule is either of the form

$$A \rightarrow a\ B$$

or of the form

$$A \rightarrow a$$

where A is a variable whose value is a nonterminal symbol and B is a variable whose value can be either a terminal or nonterminal symbol, but the a must be a terminal symbol. The initial system mentioned above where anything can become anything, is called an unrestricted rewriting system (URS). These four types of grammar form a hierarchy in terms of what they can potentially represent: the FSG is the most restricted, the CFG the next most restricted, followed by the CSG, and the URS is the most general.

One property that distinguishes between the above grammars is the self-embedding property. If a grammar has some symbol for which we can derive strings of symbols containing the same symbol flanked by other symbols, that grammar is self-embedding. If a grammar has an equivalent grammar that is not self-embedding, the first grammar is not essentially self-embedding. Equivalence here means that the set of strings of terminal symbols generated is the same (for example, with story grammars, two equivalent grammars would generate the same stories). An important theorem for formal languages states that a grammar is a FSG if and only if it is not essentially self-embedding.

An examination of the story grammars referenced above demonstrates that all except one are either a FSG or a CFG, because the left side of their rewrite rules each have only one symbol. The one exception is the Johnson and Mandler (in press) grammar. This grammar contains transformational rules that make it an URS. Also in all these grammars except one, the rewrite rules work down from abstract high level symbols until they reach an Event or Action, then the relevant rule gives a choice of rewriting the Event or Action as some high level symbol like an Episode. The one exception here is the Stein and Glenn (1979) grammar, which does not have the self-embedding property and therefore is a FSG. One

other grammar, however, comes close to not having the self-embedding property. The van Dijk (1975) grammar has only one trivial self-embedding rule:

$$\text{Event} \rightarrow \text{Event} + (\text{Event})^*$$

This rule only says that an Event can be repeated more than once in the rules where it appears. But we can create an equivalent grammar by merely substituting the right-hand side of this expression for Event whenever a single Event occurs in the other rules. In fact, the other grammars merely use the * symbol to perform this repetition function instead of having a separate rule. The new grammar created by this substitution is equivalent to the van Dijk one. Therefore the van Dijk grammar is not essentially self-embedding, so by the theorem described above this grammar is also a FSG. The rest of the grammars are essentially self-embedding and therefore are not FSGs. Thus the Stein and Glenn (1979) and Van Dijk (1975) grammars are formally FSGs; while the Rumelhart (1975), Thorndyke (1977), and Mandler and Johnson (1977) story grammars are formally CFGs; and the Johnson and Mandler (in press) grammar is formally an URS.

But what is the value of knowing these formal results? Starting with Chomsky (1957), linguists have argued that FSGs, CFGs, and CSGs are not adequate as sentence grammars. It would be surprising if a type of formalism not adequate at the sentence level were adequate at the story level. In particular, though, we can ask whether the formal arguments against FSGs and phrase structure grammars at the sentence level also hold at the story level.

First, consider the self-embedding property. Are stories self-embedding and therefore do story grammars have to be self-embedding? Self-embedding stories, with successive Subgoals creating the embedding, are actually quite frequent. For example, the *Old Farmer* story used by Rumelhart (1975), Thorndyke (1977), and Mandler and Johnson (1977) has this form. In this story, an old farmer is trying to get his donkey into the shed (overall Goal) by having his dog bark at the donkey (first Subgoal, embedded in the overall Goal), but the farmer must get the cat to scratch the dog to get the dog to bark (second Subgoal, embedded in the first), but to get the cat to scratch the dog he must get the cat some milk (third Subgoal, embedded in the second), and so on. This story structure has Subgoals embedded in each other, yet it is a perfectly acceptable story. In fact, this story is more acceptable than a sentence with this structure—e.g., "The donkey the dog the cat scratched barked at jumped into the shed." Thus the standard argument for self-embedding is stronger for stories than for sentences. The *Hen* story used by Johnson and Mandler (in press) is another example of a Subgoal embedded structure. These examples show that stories require a grammar with the self-embedding property and therefore cannot in general be generated by FSGs.

A corollary of another theorem from formal languages states that FSGs cannot generate symbol strings of the form

$$a^n b^n$$

that is, strings of a number of *a*'s followed by an equal number of *b*'s. Are there stories that fit this formula? With a few minor editing changes, we can change the *Old Farmer* and *Hen* stories into ones that fit the following formula:

$$(\text{goal})^n\ (\text{outcome})^n$$

where with a little creativity n could become any integer value. Hence once again the same argument that works against FSGs at the sentence level also works at the story level.

Thus the evidence seems to be against the adequacy of FSG story grammars, but how about CFGs? A classic linguistic demonstration of the inadequacy of CFGs involves their inability to correctly represent discontinuous constituents. This case occurs when a constituent is interrupted by another constituent (that is not part of the first), but the first continues later. For example, a story might begin with one Episode, but then another unrelated Episode interrupts it and the story returns to the first after the second ends. Consider the following story fragment:

> John was hungry so he hurried down to the local Chinese restaurant. He went inside and was looking around for a table, when his good friend David came up. David asked if John had seen the latest Bergman film. John replied that he had seen the film and had particularly liked the way the bedroom scene was shot. David then asked if John would like to go see the latest Herzog film with him that Friday. John agreed to go and David left. Then John found a table and ordered Mu Shu pork . . .

Here the film discussion Episode interrupts the Episode about the restaurant, but the restaurant episode continues when the film Episode ends. This case contrasts to the *Old Farmer* story where the interrupting and interrupted Episodes were closely related (i.e., the interrupting Subgoal is subordinate to the interrupted Goal). Hence we would want a grammar that treats those two situations differently: that is, we want one Episode interrelation rule that operates when one Episode interrupts another and the Episodes are related, but a different Episode interrelation rule that operates when one Episode interrupts another and they are not related. In other words, the rules should be sensitive to the context. But if they are, then the grammar cannot be a CFG.

A skeptic might object to our using an "artificial" story (that is, one that we wrote to make our point) in this discontinuous constituent argument. However, there are also classic literary examples of an unrelated embedded story interrupting a "top level" story that continues when the embedded story ends. Perhaps the most extreme examples are *The Arabian Nights*, *The Decameron*, and *The Canterbury Tales*. In *The Arabian Nights*, Scheherazade avoids her own murder by telling a long series of interesting stories that are unrelated to her plight. In *The Decameron*, a group of aristocrats isolate themselves in a castle to escape the plague. While waiting for the plague to pass, they amuse each other by telling stories. Similarly, the pilgrims in *The Canterbury Tales* entertain themselves on their journey by telling stories. Thus there are many "real"

examples of unrelated stories embedded in other stories and these examples support our argument against CFGs.

Let us next consider whether any phrase structure grammar, even a context sensitive one (CSG), is powerful enough to represent stories. Recall that the distinguishing characteristics of phrase structure grammars are that the rewrite rules change only one symbol at a time and that the length of the left side of a given rewrite rule is always less than or equal to the right side. Hence, a rule that deleted a story statement cannot be a phrase structure rule. But the need for deletion rules is apparent from examining even those stories cited as examples by the story grammars. These stories often omit goals and other internal states and events that are required by the story grammars. Also some of the statements in these stories occur in different orders than the orders predicted by the grammars. Reordering rules typically move several symbols at once and therefore are also not phrase structure rules. Thus stories require at least some deletion and reordering rules, so phrase structure rules (whether CFGs or CSGs) are inadequate to represent them.

When the analysis of sentence grammars reached this point, Chomsky (1965) proposed a framework that has become known in linguistics as the "Standard Theory." The proposal is that a grammar should have a base component defined in terms of phrase structure rules and a transformational component defined by URS rules that transform entire structures generated by the base component into other structures. Johnson and Mandler (in press) have proposed a story grammar of this form. This grammar has a base component defined by a set of context free phrase structure rules that are a refinement of the Mandler and Johnson (1977) grammar. This new grammar, however, also has a transformational component containing some deletion and movement (reordering) transformations. Thus this grammar fulfills the formal requirements that we have described. All of the other story grammars are either FSGs or CFGs and therefore do not meet these requirements.

This criticism would not be particularly damaging if the other grammars could remedy the situation by merely adding a few transformational rules as Johnson and Mandler have done. However, mathematical results by Peters and Ritchie regarding formal languages indicate that it is the form of the transformations that is the crucial part of the grammars (see Bach, 1974). In particular, if the transformations are chosen appropriately, the other rules in the grammar can be of almost any form. How to formally characterize the transformations required for sentence grammars is currently an unsolved problem in linguistics. If the nature of transformations is the critical part of a grammar, then the story grammar theorists have not concentrated on the essential aspect of the problem. In fact, only Johnson and Mandler (in press) have even discussed it.

One more mathematical result concerning formal languages deserves mention here. It is ironic that two developmental psychologists, Mandler and Johnson, are trying to devise a transformational grammar of stories when Wexler

and Hamburger (see Culicover, 1976) have shown that transformational grammars are unlearnable, except under a very artificial set of conditions. Hence, even if we had a fully adequate transformational grammar of stories, it would not aid developmental psychology because children would not be able to learn a grammar of that form.

We conclude from our formal analysis of stories that to be formally adequate a story grammar must be a full-scale transformational grammar. Our examination of the proposed story grammars showed that two of them were finite state grammars, two were context free grammars, but only the Johnson and Mandler (in press) grammar was a transformational grammar. Hence only the Johnson and Mandler grammar fulfills the formal requirements that will be necessary for an adequate story grammar. In the next section, we will determine the empirical adequacy of the particular transformational rules that Johnson and Mandler proposed.

EMPIRICAL EVALUATION AS GRAMMARS

In this section, we concentrate on the contents of the story grammar rules rather than their form. The evaluation technique we use here is to invent texts, then ask whether the various grammars accept them as stories. Ideally, a story grammar should be a set of rules which generates all the texts that are stories but no texts that are not stories. Thus if we find a story that a grammar does not generate, that story is empirical evidence against the grammar. In addition, if we find a nonstory text that a grammar generates, that also is empirical evidence against the grammar.

First, we note that the story grammars are not completely general, rather they apply only to stories about a single major character striving to reach a single overall goal. Thus, for example, they do not apply to stories in which the major character has multiple simultaneous (and possibly conflicting) goals (see Wilensky, 1978a and b for a discussion of such stories), nor do they apply to stories with multiple protagonists pursuing different simultaneous goals (see Wilensky, 1978b; and Bruce & Newman, 1978). But single-protagonist, single-goal stories form an important class of stories, so giving an adequate account of them would still be a major accomplishment. Therefore we will restrict our empirical evaluation to texts of this kind.

We need not look far to find acceptable stories that are not generated by the grammars. In fact, as noted in the last section, some of the stories cited as examples of the grammars will serve. Specifically, there are numerous examples of statements demanded by the grammars but missing from the story texts. There are also several examples of statements appearing in the story texts in different orders than the grammars specify. Johnson and Mandler (in press) have attempted to solve these problems by proposing deletion and movement transfor-

mations. Thus the adequacy of these transformations is crucial for our evaluation.

The transformations allow three kinds of deletions: Beginnings, Complex Reactions by characters, and Endings can all be deleted under certain conditions. We will examine the Reaction deletion rule in detail to illustrate why we think these rules are missing the real issue. The Reaction deletion rule claims that Complex Reactions (by characters) "may be deleted from any episode if the immediately prior beginning is present." However, Johnson and Mandler realize that this rule is not completely correct and qualify it by saying that it only holds when the deleted Complex Reaction is inferable from the surrounding Beginning and Attempt.

Contrasting the following two story fragments will illustrate the need for this qualification:

(A)
(1) John learned that his wife wanted a divorce. (2) John was upset. (3) He went out and got drunk.

(B)
(1) John Learned that his wife wanted a divorce. (2) John was overjoyed. (3) He went out and got drunk.

In (A) we can delete statement (2), which is the character's Reaction, because that is the Reaction readers would naturally infer from statements (1) and (3). However, the Reaction in (B) is not stereotypical, so we would not be able to delete (2) here without changing the nature of the story. Thus Johnson and Mandler's qualified Reaction deletion rule works with these examples.

However, we claim that the important element here is not the rule, but the qualification. Specifically, the syntactic categories of Beginning, Reaction, Attempt, and Ending are actually irrelevant; what is relevant is whether the candidate-for-deletion is inferable from the remaining story statements. Thus we claim that anything in a story that is inferable can be deleted, so the essential issue is determining what is inferable. But this issue is a semantic problem, so syntactic transformations will be of no assistance in its investigation.

To show that inferability is the better rule, we give an example of its application that conflicts with Johnson and Mandler's predictions. In particular, Johnson and Mandler claim that Attempts cannot be deleted, but the following example shows that Attempts can be deleted if they are inferable:

John needed a book from the library and it was soon in his possession.

In this case, obtaining the book required only executing the standard procedure of going to the library so the boring library details that would comprise this Attempt can be deleted because they are easily inferred.

Johnson and Mandler's movement transformation rules have similar problems: that is, the effective criteria for what can be moved are semantic criteria that cannot be captured by syntactic transformations. For instance, Johnson and

Mandler claim that only Goal statements can be relocated within Episodes whereas Attempt and Outcome statements cannot be moved. However, if the standard order of these statements can be easily inferred then they can appear in any order in the story text. For example, we might have the following standard Goal-Attempt-Outcome sequence:

> John wanted to be rich, so he speculated in real estate and became very wealthy.

We can move the Goal around as Johnson and Mandler claimed. The following is the same Episode in Attempt-Goal-Outcome order:

> John speculated in real estate because he wanted to be rich and he became very wealthy.

But we can also reorder the Attempt and Outcome without any problem. Here is the same Episode in Goal-Outcome-Attempt order:

> John wanted to be rich and he became very wealthy by speculating in real estate.

Thus what is essential here is not certain orderings of the syntactic categories (i.e., Goal, Attempt, Outcome), but whether the readers can infer the standard order based on their world knowledge. For simplicity, we have given very short examples, but we could lengthen any of the above elements without changing the argument. For example, we could have had a paragraph describing John's wealth in detail and then state that he obtained it by speculating in real estate.

We conclude that there are simple stories that the story grammars do not generate, even when we include syntactic transformation rules. However, an even more damaging case against the grammars is that they accept nonstories as stories. One major kind of nonstory that the grammars erroneously accept as stories is a procedural exposition. A procedural exposition is a text that instructs the reader how to do some task—that is, it communicates a procedure. Graesser (1978) proposed a grammar for procedures that is essentially the same as the story grammars. For example, the Thorndyke grammar, which we described earlier, would accept the following procedural exposition as a story:

> It is fishing season in Connecticut and a friend asks a typical suburbanite to go fishing, but the suburbanite knows that he does not know how to fish. Naturally, the suburbanite wants to impress his friend by catching some fish, so here is how he can do it. First, he needs to buy some fishing equipment. In order to do this, he goes to a sporting goods store and buys a pole, some line, some bait, and a lure. When he has this equipment, he needs to find a good place to fish. He asks his friends and looks in a guidebook . . . Having followed the principles of good fishing, the suburbanite returns home with a batch of fish.

This text (which we adapted from Graesser, 1978) contains a Setting, Theme, Plot, and Resolution as demanded by Thorndyke's story grammar. The Setting has a Time ("fishing season"), a Location ("Connecticut"), and Characters ("a typical suburbanite," "a friend"). The Theme has an Initiating Event ("a friend asks a typical suburbanite to go fishing"), a Reaction ("but he knows that he

does not know how"), and a Goal ("to impress his friend by catching a fish"). The Plot has several Episodes, the first of which has the Subgoal of buying some fishing equipment, the Attempt of going to the store and buying the various pieces of equipment, and the Outcome of having the equipment. The Ending is the Resolution of returning home with a "batch of fish."

Thus, we have demonstrated that story grammars accept as stories at least one class of nonstory text (namely, procedural expositions). We think that this discrimination failure is not accidental, because the core intuitions underlying story grammars relate to the planning or problem-solving knowledge that readers use when understanding both stories and procedural expositions. We propose that stories should be analyzed not in terms of grammars, but in terms of the knowledge that readers must use to understand the story characters' attempts to attain goals.

In addition to explaining why stories and procedures are hard to distinguish, this "planning knowledge" orientation also captures the story grammar intuitions about what makes a story "well-formed." For example, consider this story fragment:

> Little Mary heard the ice cream man coming down the street. She remembered her birthday money and rushed into the house.

Such a fragment is incomplete. As Rumelhart (1977) pointed out, the reason for this incompleteness is that the fragment does not finish describing Mary's execution of a standard plan for obtaining ice cream. This and other examples suggest that we should characterize well-formed stories in terms of characters' plans for attaining goals rather than in terms of grammars.

We conclude from our empirical analysis that the story grammars fail both of our criteria for empirical adequacy. Specifically, they fail to generate many valid stories and hence would erroneously reject them as stories. Also the grammars would incorrectly accept some nonstories as stories. We propose that investigating the knowledge people use to understand stories would be a more successful research framework than the story grammar approach. We will discuss this framework after we have evaluated the story grammars as comprehension models.

EVALUATION AS COMPREHENSION MODELS

No detailed model has been developed that describes how a story grammar would be used to understand a story, therefore we must make some assumptions about how such a model might work. Specifically, we assume that a story understander possessing a story grammar would match successive sentences of a story against the syntactic rules of the story grammar. Eventually, if the story were well-formed, some set of syntactic rules would be found that were applicable to the

text, and which describe the text's constituent structure. The semantic rules associated with these syntactic rules could then be used to build a semantic interpretation of the story. Since we need both syntactic and semantic relations for a comprehension model, we will use the Rumelhart (1975) story grammar instead of the Thorndyke one. We choose Rumelhart because that grammar has the most elaborate group of semantic relations.

Even this very general model of how a story grammar might aid an understander does not hold up under closer scrutiny. For example, consider the following story given in Rumelhart (1975):

> (1) Margie was holding tightly to the string of her beautiful new balloon. (2) Suddenly, a gust of wind caught it (3) and carried it into a tree. (4) It hit a branch (5) and burst. (6) [Margie was sad.] (7) Margie cried and cried.
>
> (The brackets in (6) mean that the event was an inference, not an actual part of the story text).

The causal connection between (5) and (7) is an INITIATE link, which is associated with RULE 3 of the Rumelhart story grammar:

> RULE 3: **Episode** → Event + Reaction
> ⇒ INITIATE(Event,Reaction)

Here the semantic part of the rule is given on the first line and its semantic counterpart is given on the second line after the double arrow. Episode is a syntactic structure that may consist of the constituents Event and Reaction, in which case the meaning of the Episode is that the Event INITIATE'd the Reaction. A Reaction is further defined as

> RULE 5: Reaction → Internal Response + Overt Response
> ⇒ MOTIVATE(Internal Response,Overt Response)

and each of these is defined further in the grammar.

Now suppose a reader were trying to understand the Margie story and came across sentences (5) and (7). Presumably, the reader would interrogate its story grammar to find the syntactic rules applicable at this point in the story, and would eventually find RULE 3. To determine whether RULE 3 were applicable, the reader must check to see if sentences (5) and (7) fit the constituent structure as defined by the right hand of this rule. This would require determining whether sentence (7) were of the form Reaction. Reaction would then be further broken down by RULE 5, and the reader would have to decide whether an Internal Response and an Overt Response were contained in the text. Let us ignore for the time being the difficulty presented by the fact that the Internal Response is not explicit in the text but had to be inferred. What we are left with is the problem of whether (7) can be an Overt Response. Checking this seems simple—the reader merely expands Overt Response and finds that RULE 7 allows an Overt Response to go to an Action, which is not further defined by the grammar. The reader then asks whether (7) is an Action that could have been MOTIVATE'd by an Internal Re-

sponse INITIATE'd by Event (5). Some semantic routine verifies that indeed this is the case. The reader has thereby found the syntactic rules that account for this segment of the story (e.g., RULE 3 and RULE 5), and the semantic counterparts of these rules could now be used to build an interpretation of the story.

The problem with this argument is that in order to determine the applicability of a syntactic rule, the reader must have first determined the semantic relationship between (5) and (7). It was necessary to determine this semantic relationship because the syntactic category Action did not provide enough to decide whether (7) could be an Overt Response to (5). For example, suppose sentence (7) were changed to

(7)′ The broken balloon fell to the ground.

In order to determine that this is not an Action that is an Overt Response to (5), a reader would have to examine the semantic relationship between the two sentences, since both (7) and (7)′ appear to be indistinguishable by syntatic considerations alone. So in order to apply a syntactic rule to a pair of sentences, a reader must first determine if the rule is applicable. For this, the reader must determine the semantic relationship between the two sentences. But it was the semantic relationship between the two sentences that we were hoping the syntactic rule would lead us to. Hence before a rule can be applied, the reader must have already accomplished the purpose for which the rule was designed.

But perhaps the syntactic rules would at least be useful in suggesting which semantic relationships to test. In order to determine which syntactic rules to try, we must know the syntactic class of each proposition. But as the above example demonstrated, the syntactic class of a proposition is dependent upon the semantic relationship of the proposition to other propositions in the text—a Reaction is a Reaction only if it could be semantically related to an INITIATE'ing event, not because of some criterion inherent in that proposition. Since it is not known in advance which semantic relationship one sentence in a text will have to the next sentence in the text, we would need a syntactic rule corresponding to each semantic relation. In fact, these are just the kind of rules Rumelhart provides. Thus there would be as many syntactic rules as there are conceivable semantic relationships between sentences. Testing each rule amounts to enumerating all possible semantic relationships between the sentences, trying each one, and choosing its syntactic counterpart. But, of course, it is the enumeration of semantic possibilities that we were hoping the syntactic rules would help to circumvent. In any case, it is only possible to arrive at the syntactic interpretation *after* the semantic relationship has been found.

The principal reason why these syntactic rules are not particularly helpful in story understanding is that the syntactic classes they presuppose cannot be defined independently of the semantic relationship between the sentences. One cannot look at a proposition itself and assign it to a syntactic category. A careful examination of Rumelhart's rules will reveal that each syntactic class is really

just a name for a proposition participating in a particular semantic relationship. For example, an Overt Response is a syntactic category, but membership in it can only be decided if the Overt Response action can be semantically related to an INITIATE'ing Event and a MOTIVATE'ing Internal Response.

It might still be argued that the syntactic rules would be of some use because they do rule out some possibilities. For example, State and Event are syntactic categories whose membership may be determined by examining each proposition independently. Since there are inherently different semantic relationships between a State and an Event than between two Events, recognizing this syntactic feature would limit the number of semantic interpretations which an understander would need to try. The problem with this argument is that it trivializes the notion of a story grammar; the claim amounts to saying only that there are rules of causation—e.g., events are related causally to other events differently from how events are related causally to states. While other natural language researchers have postulated such rules (e.g., see Schank, 1975 and Wilks, 1977), the claim has no direct bearing on the notion of story. That is, rules of causation are rules about how conceptualizations relate to each other, not rules about where to find two sentences in a story that are likely candidates for causal connection. If the only syntactic categories that can be determined before looking at causal connections are those that correspond to some conceptual category (like State, or Action), then the only thing the story grammar can say about the structure of the story is that a sentence in a story may be related to another sentence in a story by an appropriate causal link. Since all the syntactic categories that are story-specific (such as Reaction and Overt Response) cannot be determined prior to semantic analysis, we reduce the notion of the "structure" of the story to the notion that the events of the story may be causally related.

To make matters worse, the classification of sentences into syntactic categories like State and Action does not completely determine the kinds of causal links between sentences because the connection between sentences may involve a causal chain containing events *inferred* by the story understander. This has serious implications for the ability of a story grammar to aid comprehension. Suppose we were trying to find a story grammar rule that would account for the connection between (5) and (7) of the Margie story. Since none of the rules in our story grammar apply directly to the actual sentences in the story, some inference mechanism must be used to find a plausible connection between the two sentences. Inference procedures are based upon the semantics of the story, and therefore must operate independently of the syntactic considerations suggested by Rumelhart's rules. For example, that Margie was sad because her balloon broke and that her sadness made her cry must be inferred based on our knowledge about children and how they may react to loss of possessions. Syntactically, there is nothing to suggest that the relationship between (5) and (7) should be different from the relationship between (4) and (5). It is obvious to most readers that the balloon did not undergo an Internal Response upon striking the tree, which then

MOTIVATE'd its Overt Response of bursting. Likewise, it is unlikely that the balloon bursting CAUSE'd Margie to cry in the same way that the balloon striking the tree CAUSE'd the balloon to burst. But these relations are implausible because of the nature of the content of these events—e.g., balloons don't have internal responses; people do. Thus the appropriate semantic relationship between sentences becomes evident only after an inference process has found a way to relate them.

Once again, we are caught in a quandary. To apply syntactic grammar rules, we must first know what the events are that constitute the story. Finding these events may require the use of inference procedures to postulate implicit events. These inference procedures can determine that an event needs to be inferred based on knowledge about the possible semantic relationships among events. Thus it is only after these inference procedures have run, unguided by syntactic considerations, that we can hypothesize the actual propositions to which the syntactic story grammar rules are to apply. But by this time the inference procedure has already determined the semantic relationships among events. This was the task we assumed would be aided by story grammar rules. So the argument that a set of syntactic story grammar rules would help a reader to understand a story is circular. In order to determine the constituent structure of the story, we need to first have understood the story. But in that case, analyzing it into its consituent structure becomes unnecessary.

CONCLUSION

We have evaluated story grammars from three perspectives. The evaluations indicated that the Johnson and Mandler (in press) grammar is the only one adequate from a formal perspective and all the grammars have major deficiencies when examined empirically as grammars or when evaluated as a basis for a model of story comprehension. These deficiencies lead us to conclude that the story grammar approach to investigating story understanding is not a promising one. The empirical-grammar and the comprehension evaluations provided several examples where it was more useful to focus on the content of the story rather than the syntactic structure emphasized by the grammars. In particular, the importance of the semantic processes was evident during both of these evaluations. Hence a better research orientation would be to focus on the kinds of knowledge needed to understand story content and on how that knowledge is used during understanding.

The important question from this point of view is distinguishing between texts that are understandable and those that are not. In contrast, the story grammars try to determine whether a text is a structurally "well-formed" story. As we have argued above, determining the "well-formedness" or grammaticality of a story presupposes understanding the story. Since the purpose

of the grammatical structure of a story is to aid in understanding the story, there is no reason to determine the structure because we must have understood the story before we can discover the structure. Hence the important issue for investigation is the nature of understanding, not grammaticality.

The approach we advocate is to explore the kinds of knowledge that people use in story understanding. From this perspective, it is desirable to discover knowledge structures that are needed for understanding several different kinds of text. For example, as we pointed out above, "planning" knowledge is used in understanding both stories and procedural expositions. Strangely, from the story grammar perspective such a discovery is a disadvantage because it decreases our ability to discriminate stories from nonstories. Hence the story grammar framework focuses on spurious issues while obscuring important ones.

Several investigators have recently taken the approach we advocate for studying story understanding: for example, Schank and Abelson (1977) have proposed several kinds of knowledge needed to understand story content; Cullingford (1978) and Bower, Black and Turner (1979) have explored the use of stereotypical action sequences (scripts) in story comprehension and memory; and Wilensky (1978a and b) and Black and Bower (in press) have investigated the use of mundane planning knowledge in story comprehension and memory. These content oriented approaches are examining the important issues in story understanding, whereas the story grammars are not.

ACKNOWLEDGMENTS

We thank Gordon Bower, Herbert Clark, Jean Mandler, Roger Schank, Chris Riesbeck, and Jaime Carbonell for comments on earlier versions of this paper.

REFERENCES

Bach, E. *Syntactic theory*. New York: Holt, Rinehart and Winston, 1974.

Black, J. B. Story Memory Structure. Manuscript submitted for publication, 1979.

Black, J. B. & Bower, G. H. Story understanding as problem-solving. *Poetics*, in press.

Bower, G. H., Black, J. B., & Turner, T. J. Scripts in memory for text. *Cognitive Psychology*, 1979, *11*, 177–220.

Bruce, B. & Newman, D. Interacting plans. *Cognitive Science*, 1978, *2*, 195–233.

Chomsky, N. *Syntactic structures*. The Hague: Mouton, 1957.

Chomsky, N. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press, 1965.

Colby, B. N. A partial grammar of Eskimo folktales. *American Anthropologist*, 1973, *75*, 645–662.

Culicover, P. W. *Syntax*. New York: Academic Press, 1976.

Cullingford, R. E. *Script application: Computer understanding of newspaper stories*. Research Report No. 116, Department of Computer Science, Yale University, New Haven, CT, 1978.

Graesser, A. C. How to catch a fish: The memory and representation of common procedures. *Discourse Processes*, 1978, *1*, 72–89.

Hopcroft, J. E. & Ullman, J. D. *Formal languages and their relation to automata*. Reading, MA: Addison-Wesley, 1969.

Johnson, N. S. & Mandler, J. M. A tale of two structures: Underlying and surface forms in stories. *Poetics*, in press.

Lakoff, G. P. Structural complexity in fairy tales. *The study of man*, 1972, *1*, 128–190.

Mandler, J. M. & Johnson, N. S. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 1977, *9*, 111–191.

Propp, V. *Morphology of the folktale*. Austin: University of Texas Press, 1968.

Rumelhart, D. E. Notes on a schema for stories. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975.

Rumelhart, D. E. Understanding and summarizing brief stories. In D. La Berge & J. Samuels (Eds.), *Basic processes in reading comprehension*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.

Schank, R. C. The structure of episodes in memory. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975.

Schank, R. C. & Abelson, R. P. *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.

Stein, N. L. & Glenn, C. G. An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *New directions in discourse processing*. Norwood, NJ: Ablex, 1979.

Thorndyke, P. W. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 1977, *9*, 77–110.

Van Dijk, T. A. Recalling and summarizing complex discourse. Unpublished manuscript, University of Amsterdam, Amsterdam, The Netherlands, 1975.

Wall, R. *Introduction to mathematical linguistics*. Englewood Cliffs, NJ: Prentice-Hall, 1972.

Wilks, Y. What sort of taxonomy of causation do we need for language understanding? *Cognitive Science*, 1977, *1*, 235–264.

Wilensky, R. Why John married Mary: Understanding stories involving recurring goals. *Cognitive Science*, 1978, *2* 235–266. (a)

Wilensky, R. *Understanding goal-based stories*. Research Report No. 140. Department of Computer Science, Yale University, New York, CT, 1978. (b)