ILLINOIS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

\. /5 2/

? 6/

2//-//7 9

T
E **R**
C **E**
H **P**
N **O**
I **R**
C **T**
A **S**
L

Technical Report No. 168

PROBLEMS AND TECHNIQUES OF TEXT ANALYSIS

G. M. Green, R. N. Kantor, J. L. Morgan,
N. L. Stein, G. Hermon, R. Salzillo,
and M. B. Sellner
University of Illinois at Urbana-Champaign

B. C. Bruce, D. Gentner, and B. L. Webber
Bolt Beranek and Newman Inc.

April 1980

Center for the Study of Reading

UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN
51 Gerty Drive
Champaign, Illinois 61820

BOLT BERANEK AND NEWMAN INC.
50 Moulton Street
Cambridge, Massachusetts 02138

CENTER FOR THE STUDY OF READING

Technical Report No. 168

PROBLEMS AND TECHNIQUES OF TEXT ANALYSIS

G. M. Green, R. N. Kantor, J. L. Morgan,
N. L. Stein, G. Hermon, R. Salzillo,
and M. B. Sellner
University of Illinois at Urbana-Champaign

B. C. Bruce, D. Gentner, and B. L. Webber
Bolt Beranek and Newman Inc.

April 1980

University of Illinois
  at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge, Massachusetts 02138

This page is intentionally blank.

Table of Contents

List of Figures and Tables

Problems and Techniques of Text Analysis

I. The Text Analysis

    The goal of the Text Analysis groups of the Center for the Study of Reading is to investigate the problem of reading comprehension from the standpoint of properties of texts, in order to understand the contribution of texts to difficulty or ease of reading, and to construct appropriate theories to account for text properties. The need for this kind of research is clear. Given a solid understanding of these matters, texts could be more effectively evaluated for quality, and for level of difficulty; this would allow careful matching of the text to the needs of the child, and it would probably be possible to construct more accurate diagnostic tests of reading proficiency. A more complete understanding of the text properties and their role in reading comprehension would no doubt serve as a basis for instruction, as well.

    It is important to keep in mind, of course, that difficulty is not necessarily a bad thing. A gross method of detecting difficulty level would be nearly worthless if it could not distinguish between a text that was well-constructed but made challenging demands on the reader, and one that was so poorly constructed that it contained unnecessary obstacles to comprehension. Children, like adults, can learn by being challenged; but a poorly constructed text can not only obstruct learning but can confuse or discourage a less-than-proficient reader. It is by no means obvious just what it is that makes a text a challenge or an obstacle. We have seen texts that had been modified for poor readers by abridging and modifying

longer texts, that impressed us as harder, not easier, than the original.
We suspect that these are not isolated instances. For example, it is likely
that in forcing a complex story or exposition onto the Procrustean bed of
small vocabulary, short sentence length, and short text length, one requires
the reader to make more complex inferences to reconstruct the intended in-
formation structure of the text. Consequently, it is essential to develop
tools not only for evaluating the difficulty level, but also for distin-
guishing challenging from confusing texts.

Ideally, then, it would be possible to examine a given text using
simple replicable methods of evaluation to determine not only the level of
difficulty, but in fact the sources of difficulty, in a given text. But
such methods simply do not now exist. The state of the art in such matters,
in particular theoretical work on text properties in linguistics, psychology,
and artificial intelligence, just has not reached a state that affords any
kind of non-subjective methods of analysis for any but the simplest kind of
text properties, like lexical frequency counts and measures of sentence
length.

The main purpose of the research of the Text Analysis Group is to
advance the state of theoretical work on text properties. This research
bears on that goal in three ways: first, it allows the sharpening of the
tools of analysis, and a determination of which methods of analysis are
likely to be usable and fruitful, albeit subjective and intuitive. As an
example, we have not included given-new or topic-comment analyses of the
texts, since the methods of analysis we have attempted to use turn out to
be either impossible to apply to non-trivial texts, or when applied

consistently, lead to intuitively absurd results. Such an outcome shows the poverty of the theories with which such methods are associated, and calls for further theoretical work on these matters.

Second, this work is exploratory, in that close analysis of texts for several kinds of properties is likely to reveal new classes of phenomena, and to suggest research on their role in text comprehension.

Third, we are convinced that the state of theoretical work on text properties can only be advanced beyond its present state by consideration of data bases considerably richer than presently available. Most theoretical discussions of text properties are based on single texts, often of trivial size. We intend to remedy the situation by collecting a significant body of analyzed texts, with which to compare developing theories of text properties. The research in this report is a step toward that goal. This kind of analysis will be continued, and we are developing systems for annotating, storing, and retrieving analyses of texts.

As a consequence of the exploratory nature of the work in this report, parts of it are uninterpreted at present. For example, one cannot know what to make of the section on the relation between syntactic and line-end properties until similar analyses can be compiled for a number of texts, and experimental work is done to determine the effects of the differences.

II. Lexical Analysis

A. Statistical Lexical Analysis

The statistical analysis consists of a word frequency count of the words in the text. The first part of this section describes the way the frequency counts were calculated. The second part of this section presents

an alternative to statistical analyses which treat each <u>orthographic word</u> as a separate unit. This analysis suggests that a text may be analyzed by <u>semantic lexical units,</u> which are semantic units that the reader must comprehend, and which have the property of being psychologically real to the reader of a text. This section also contains discussion of a number of questions concerning the utility of simple word frequency counts.

Statistical lexical analysis requires decisions about what counts as one word (e.g., are contracted forms like <u>don't</u> one word or two?) and about what counts as instances of the same word for the purposes of frequency counts (e.g., are <u>taxi</u> and <u>taxis</u> the same word?). In our analyses, contractions were treated as containing two words, and words related by inflectional morphology (e.g., <u>taxi</u> and <u>taxis</u>; <u>be</u>, <u>am</u>, <u>is</u>, <u>are</u>, <u>was</u>, and <u>were</u>) were treated as instances of the same word.

In the frequency counts, each sequence of letters delimited by either spaces or apostrophe was treated as a separate word. That is, no matter what the semantic relationship of a lexical item to other words in the sentence, it was considered independently of any semantic unit containing it. For example, the word <u>for</u> was separated from <u>waiting for</u>, a semantic unit in which it is contained, and counted as a separate lexical item.

This section treats the utility of viewing the discourse as consisting of more semantically integrated lexical units. There is no known replicable procedure for dividing the sentences into such units--linguists debate such issues constantly. Nonetheless, it seems that certain kinds of semantic analyses of the words in the discourse may have some relevance to the study of difficulty of texts.

Our primary concern is that frequency counts of words are counts of

form. If frequency counts and counts of average number of words per sen-

tence do correlate with difficulty of a text, these statistical measures

must only be symptomatic of deeper properties of texts. Below we discuss

some examples from BABAR and DESERT[1] to show the kinds of considerations

that might go into a semantic lexical analysis. In the course of this dis-

cussion, some of the problems are pointed out which are connected with

simple word form frequency counts.

One interesting possibility which the semantic lexical analysis affords

that the simple word count does not is that of lumping together units that

have the same reference. So, for pronouns, it is possible to separate the

he's that refer, for example, to Babar, e.g.,

93-94   But Babar can't eat. He is thinking about his crown,

from those he's and him's that refer, for example, to Arthur and the

Mustache-man, e.g.,

58-63   Arthur climbs out of the water just as fast as he can. He
        is all wet. He sees the man with the bag. Arthur calls to
        him. But the man does not hear him. He is walking away.

It is then possible to count separately the different kinds of reference

expressions used for the respective referents. Perhaps there will be cor-

relations between the number of different types of reference to the same

referent and the difficulty of the material under study. That is, it may

be useful to look at the various lexical descriptions, pronominal references,

various instances of definite reference, e.g., the man, the man with Babar's

bag, the Mustache-man, and references with demonstratives, e.g., that

Mustache-man. A word frequency count will not give us this kind of informa-

tion.

Word frequncy counts also make no distinction between content words

and homophonous words that serve strictly grammatical functions. For

example, in

32 Do you think that all deserts are sandy and flat?

from DESERT, the word that is a subordination marker, serving a syntactic

function, and should be identified as such. Further, it should not be

confused with demonstrative that as in BABAR

33 "We'll find that Mustache-man,"

where the demonstrative does have semantic content.

Word frequency counts likewise confound the so-called "weather it," as

in DESERT,

8 In the desert it may not rain for five years,

where the phrase it...rain may well be taken as a single unit, with anaphoric

non-referential it's as in

24-25 A man lost in the desert knows his camel will help him. It
        will find a water hole.

and anaphoric referential it's, as in (from BABAR)

30-31 "I need my crown!" says Babar. "I must wear it tonight!"

It is not obvious that lumping these it's together is justified.

There is still a great need for linguistic research to develop reliable

criteria for isolating multi-word semantic units. With such criteria it

would be possible to perform more refined statistical analyses, and to

determine, for example, whether writers or texts vary significantly in the

degree of coincidence between orthographic and semantic units, and how this affects ease of comprehension.

B. Semantic Lexical Analysis

The opening sequence of a book in a beginning reading series, Babar Loses his Crown, tells about the Babar family travelling to Paris.

The Babar family is going to Paris....
Now the Babar family is on the train.
The train is coming into Paris....
Now they are off the train
with all their bags.
The Babar family is waiting
for a taxi.
The taxi takes them to their hotel.
Celeste and the children walk inside....
In their hotel room....

A fair amount of space is devoted to the journey. The trip could have been summed up more briefly, as in

The Babar family took a train to Paris, taxied to their hotel and
went inside. In their hotel room...

or even

The Babar family travelled to Paris. In their hotel room,...

The sequence of locations, pathways, and methods of transportation can be presented in a detailed fashion, with the states and events spelled out, or can be abridged and compressed into one verb (e.g., travelled). How do these choices affect comprehension?

Stories tell about happenings--about actions, events, processes, states, relationships. In this section we discuss the semantics of verbs, which convey much of this information. First, we describe a notational system in which verb meaning distinctions can be expressed and discuss its psycho-logical interpretation. Then, this system is applied to the sample text

above. Next, we discuss the ways in which verb semantic structure may

affect comprehension, and describe some research in the area. Finally, we

suggest some lines along which further research might proceed.

1. Representation of Verb Meaning

There are many formal systems for representation of verb meaning. These

models of verb meaning differ from one another in detail, but there is wide-

spread agreement on the idea that verb meanings can be represented in terms

of interrelated sets of subpredicates, such as CAUSE or CHANGE. These sub-

predicates express the inferences that are normally made when the verb is

used. These inferences are highly interrelated, and the representation of

verb meaning must indicate the relationships among the sets of subpredicates.

The notation used here is a network format developed by the LNR Research

Group. In this representation system, both verb meanings and events are

expressed in the same terms--as states, changes of state, actionals,

etc.

2. The Elements of Verb Meaning

When people talk about happenings in the world, they distinguish several

types of conceptual possibilities. The simplest kind of relationship is the

state. A stative predicate conveys a relationship that endures in time

between two arguments, normally an object (or person) and a value within

the conceptual field specified by the stative. For example, consider the

sentence shown in Figure 1.
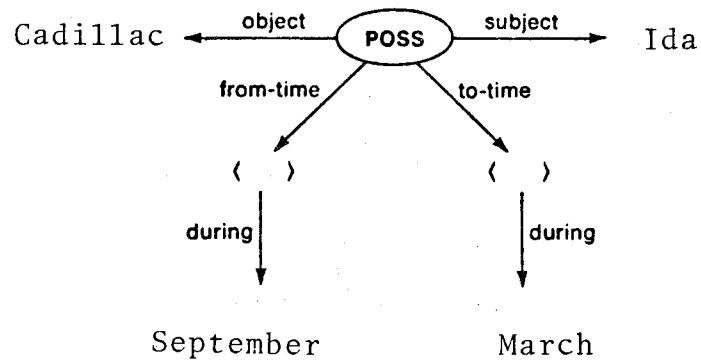
Figure 1

Ida owned a Cadillac from September
to March.

The verb <u>own</u> conveys that a relationship of possession existed between Ida and the Cadillac for some duration. Here the state is one of possession. A large number of such states, including location (<u>to be at</u>, <u>to remain at</u>, etc.) and various emotions (<u>to hate</u>, <u>to love</u>, etc.) can appear in verb meaning. Stative concepts are expressed not only by verbs, but also by prepositions and adjectives.

In addition to simple stative relationships, verbs, unlike adjectives and prepositions, can be used to convey changes of state. Following Chafe (1970), we will refer to this kind of change of state as a process. For example, the sentence

Sam received a rose.

tells us that a change of possession occurred such that an earlier state in which a rose was possessed by some unknown person changed to a state in which Sam possessed the rose. This is shown in Figure 2.
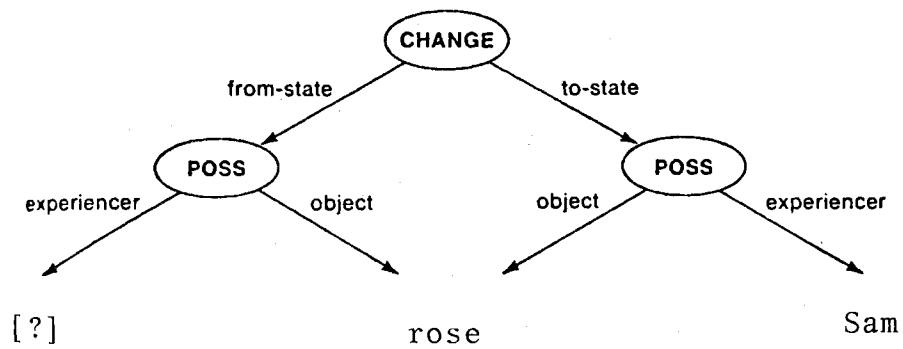
Figure 2

Sam received a rose.

More commonly, verbs express not simple changes of state but causal changes of state. An agent may cause a change of state that relates to another object, or the agent himself (or herself) may be the experiencer of the change of state. The locational verb move can be used in either way, as in the following examples:

a. Ida moved the car.

b. Ida moved to the front seat.

The representations of these sentences are given in Figures 3a and 3b.
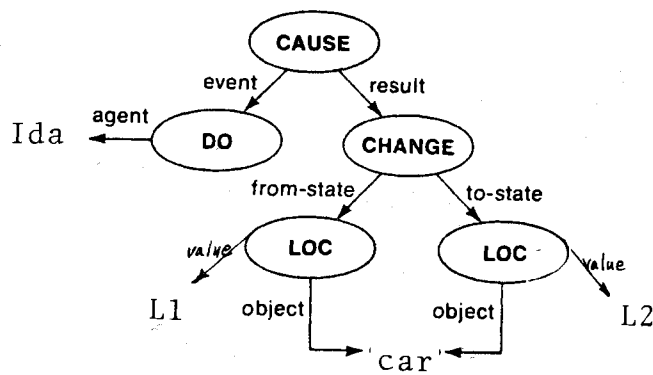
Figure 3a

Ida moved the car.
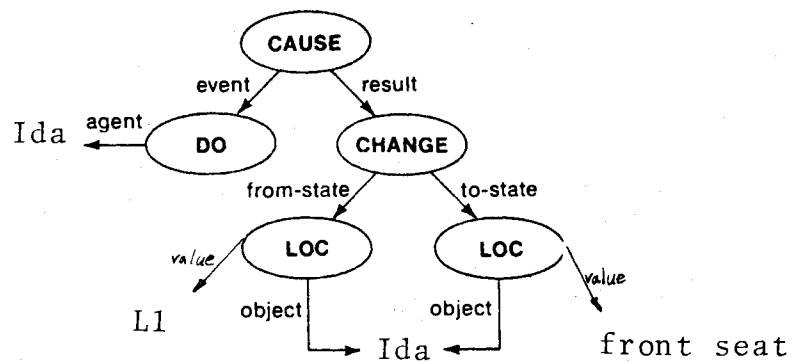
Figure 3b

Ida moved to the front seat.

Notice that in both these cases the precise activity engaged in by Ida is unspecified. However, there are also a great many English verbs in which the causal action is partially or wholly specified, for example, among the location verbs: walk, saunter, meander, stride, run, sprint, race, trot, job. (See Miller (1972) for a more extensive discussion of the verbs of location.) At present we know of no formal representational nota-tion capable of dealing with physical actions. Miller (1972) and Miller and Johnson-Laird (1976) have discussed the verbs of motion, and Greeno and Gentner have developed some tentative representations of some verbs

of mixing (Gentner, 1978). But neither of these attempts goes far enough to capture the richness of physical knowledge that we use in distinguishing, for example, _sauntering_ from _striding_. Figure 4 shows the representatives of the verbs _mix_ and _stir_ developed by Greeno and Gentner.

14
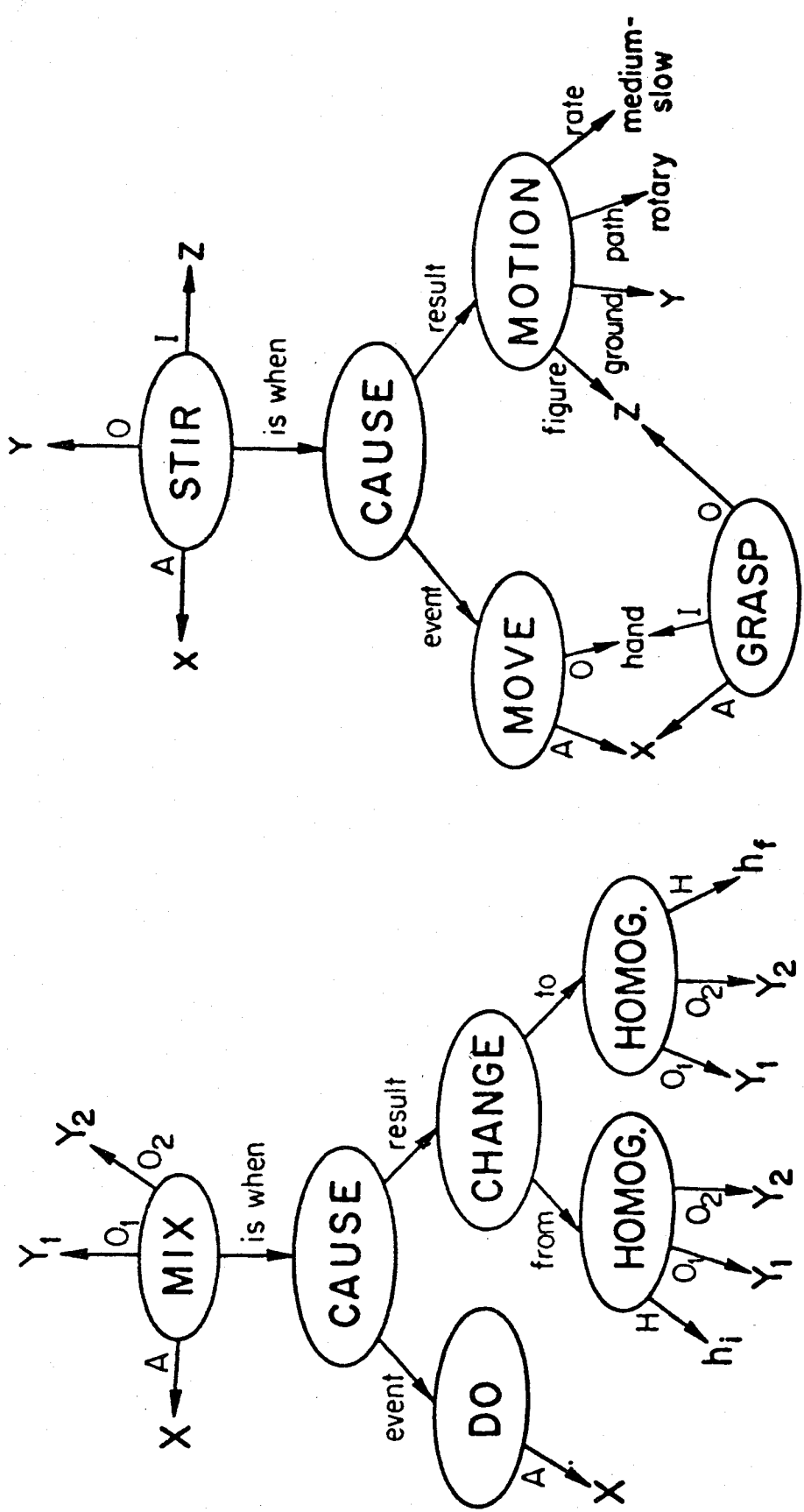
Figure 4

Representations of the meanings of
mix and stir.

15

Stir is a pure actional verb. There is no specification of the change-of-state that should occur. Only the action is specified. Stirring may be done in order to change homogeneity, as when stirring sugar in coffee, or it may be done to change temperature, as when stirring too-hot black coffee, or it may be done to prevent something from burning. Mixing is just the opposite. Here the process (change of state) is specified to be a change of homogeneity. The action by which this change is produced can be stirring, shaking, using a blender, or whatever else works.

An event can be a change of state, a causal change of state, an action, or a concatenation of events. A detailed discussion of the notions of states and events as they are used in the LNR system is given by Rumelhart and Norman (1975). Further discussion of verb semantics can be found in the articles by Abrahamson (1975), Gentner (1975), Munro (1975), and Rumelhart and Levin (1975) in the same volume.

3. Psychological Implications

The intention in writing out verb representations is to capture the set of immediate inferences that people normally make when they hear or read a sentence containing the verb. The system is decompositional, in that it is assumed that these networks of meaning components are substituted for the verbs during comprehension. This substitution process is assumed to be immediate and largely automatic, and it is assumed that the set of components associated with a given word is reasonably stable across tasks and contexts.

The representations should satisfy some psychological criteria. First, they should accurately capture people's internal meanings. Thus, we should find that the representations agree with people's intuitive notions of synonymity and similarity in meaning (see Rumelhart and Norman, 1975). One measure of this overlap is the likelihood that people will confuse words in memory. In an experiment in sentence memory, using verbs of varying semantic overlap, Gentner found that subjects did indeed confuse the verbs in exactly the way predicted by the theory (Gentner, 1975). The correlation between the number of confusions subjects made between two verbs and the semantic overlap between the verbs, as predicted from the representations, was quite high.

In fact, the correlation between representational overlap and number of confusions was slightly higher, though not significantly so, than the correlation between the number of confusions and the rated similarity between the verbs (the similarity ratings were generated by a different set of subjects).

Another psychological criterion is that the representations should be at the correct level of completeness. That is, the representational structure for a given verb should show the almost-inevitable inferences that are made when a verb is used, but should not show extremely rare inferences that can be made in only a few contexts. These must be derived from the interaction between the representation of the verb and that of the contexts. This brings us to another requirement.

Another important psychological requirement is upward-compatibility. The basic notions of state, change of state, cause, and so on must be

combinable into networks larger than the individual sentence. The infer-
ences derived from words decomposing into semantic components must inter-
act with information from the context. Sometimes some of the bottom-up
inferences, represented by subpredicates, will be overridden by top-down
contextual information, as in metaphorical uses of verbs. But, although
high-level inferences and context-based expectations are undoubtedly an
important part of meaning processing, still this top-down knowledge must
itself be based in part on the bottom-up knowledge derived from the indi-
vidual word meanings. Ultimately, it will be necessary to model word
meanings, high-level structures, and the processes that bridge between
them. In the examples in this paper, we use only fairly simple rules of
combination to join sentence representations.

Another psychological requirement is that our representations should
be able to capture the way in which word meanings are learned. Indeed,
children's acquisition of the verbs of possession fits very nicely with the
predictions generated from the model, if one makes the Clark (1973) assump-
tion of gradual acquisition of semantic components (Gentner, 1975).

With these psychological issues in mind, we can then discuss the means
by which a text achieves the desired structure, and consider some of the
issues relevant to comprehension.

4. Semantic Structure and Comprehensibility

a. Semantic Function

A fundamental distinction is whether a word is used as a predicate
(e.g., a verb) or an argument (e.g., a concrete noun). There are two
opposing lines of argument as to which of these functions most affects text

understanding. Proper nouns and concrete nouns, the quintessential argument terms, are relatively inert semantically. They function as pointers to the world, but their semantic structures perform no connective function. In contrast, the semantic structure of a verb serves to connect the noun concepts into a unified proposition. By this line of reasoning, failure to comprehend the verb should therefore lead to more confusion and more disturbance in overall understanding of the text than failure to comprehend a noun. In terms of the schematic diagrams shown in the figures, failure to comprehend a noun means that the language user must store an empty or incorrect node in one of the argument slots. Failure to comprehend the verb could leave the language user with an unorganized heap of nouns that cannot be connected with one another or integrated with the rest of the text.

Preliminary evidence indicates that verb meanings are altered more than noun meanings when subjects try to make sense out of bizarre sentences like The lizard worshipped. Further investigation might inlcude protocol analysis of what happens when children encounter unknown or only partially known words in various semantic roles. The results might indicate what kinds of words in a text must be taught most carefully in order to avoid serious confusion.

b. Case Structure

A clear distinction among verbs is how many arguments they take. Some transitive verbs, such as sell, can take four noun arguments (the buyer, the seller, the object, and the payment) in addition to the time-of-occurrence. At the other extreme, stative verbs take only two noun

arguments (an object and a value) along with the duration (initial and final times). One issue to investigate is whether a greater number of many-place predicates in a text leads to more difficulty in comprehension.

c. Semantic Complexity

Another relevant variable is semantic complexity: the number of under-lying components that make up the basic meaning of a word. More complex meanings correspond to more specific references to objects or events. For example, stride is more specific than go. Its meaning contains more sub-predicates. We know more having heard sentence (a) than sentence (b).

(a) Ida strode across the field.

(b) Ida went across the field.

Various researchers have proposed that semantic complexity may affect com-prehensibility, generally on the assumption that more complex semantic structures are harder to process. However, studies performed on adults have shown no evidence that more complex words lead either to longer re-action times or to greater processing loads than do simpler words (Kintsch, 1974).

There may be more relationship between complexity and difficulty in children than has been found in adults. Research in child language has shown that young children often fail to comprehend the full meanings of complex terms (e.g., Bowerman, 1975; Clark, 1973; Gentner, 1975). Working with the verbs of possession, we have observed that children act out the simple verbs give and take correctly before they act out the complex verbs sell and buy. Further, when young children (around 5 years) are asked to act out sell (as in "Make Ernie sell Bert a boat."), they act

out give instead (a boat is transferred from Ernie to Bert); the same is true for buy and take. Our interpretation, consistent with Clark's (1973) semantic features analysis, is that children correctly act out the parts of a word's meaning that they are conceptually able to comprehend. At any given time, the child uses only the components that he has so far acquired in comprehending language.

This being the case, the next question is what effect the use of buy and sell rather than take and give has on children's comprehension of text. The young child, having read that Ernie bought a car, may understand that Ernie acquired possession of a car, but not that he did so by means of an agreement with someone else by which Ernie was obligated to give some money in return for the car. Thus, it seems inevitable that comprehension of the immediate inferences intended in the sentence will be affected. We can then ask how this local lack of complete comprehension of buy will affect the overall comprehension of the story. Clearly two extreme cases need to be distinguished: those in which the contractual notions involved in buying are important to the sense of the whole text and those in which they are not. In the former case, failure to comprehend those parts of the meaning of buy will certainly have a pervasive effect on comprehension. In the latter case, there is at least a chance that children can use those parts of the meaning of buy that they understand and leave the rest aside without its affecting their overall comprehension of the text.

One avenue of research that this suggests would be to examine children's comprehension of text that includes either simple or complex word meanings. These complexity variations should occur either in the noun

position or in the verb position. Table 1 (following page) outlines an experiment designed to investigate these issues on several levels and with various measures.

d. Semantic Complexity and Semantic Connectivity

So far we have considered the possibility that an increase in difficulty of comprehension might be caused by use of semantically complex terms. On the other hand, the additional semantic components in a complex verb may set up additional connections among the nouns in the sentence. In this case, the more complex verb could lead to a richer and more highly interwoven text representation, and thus to better memory for the text. Gentner found this kind of improvement in connectivity in a series of experiments in sentence memory (Gentner, 1978). Subjects were given sentences that differed in the semantic complexity of their verbs, such as the following pair of sentences:

(simple)   Ida gave her tenants a clock.

(complex)   Ida sold her neighbor some art posters.

They were better able to recall the recipient nouns (tenants and neighbors) when a complex connective verb such as sell was used than when a simple verb such as give was used. (Object-verb pairings were, of course, counterbalanced.)

It is crucial for this effect that the additional semantic information in the complex verb is such as to connect the nouns in question. This is clearly true in the case of sell versus give, as can be seen in Figures 5a and 5b.

Table 1

An experiment to test the effects of verb semantic structure on reading

Purpose: to see whether semantic complexity and connectivity have effects
          on high-level reading performance.

Independent Variables	Levels
Semantic complexity	semantically complex words versus semantically simple words
Semantic function	predicate (verb) versus argument (simple noun)
Semantic connectivity	connective complexity in predicates versus nonconnective complexity
Amount of high-level structure	strong story constraints versus loose story constraints
Reader expertise	poor readers versus good readers at same age
	younger children versus older children

Dependent Variables	
Ease of initial comprehension	time to read story errors in reading out loud accuracy in immediate questioning
Later memory	ability to retell story[a] ability to recognize parts of the story

NOTES

    [a]Memory for material will be assessed both by verbation accuracy and
by correctness of content.

23

Ida gave her tenants a clock
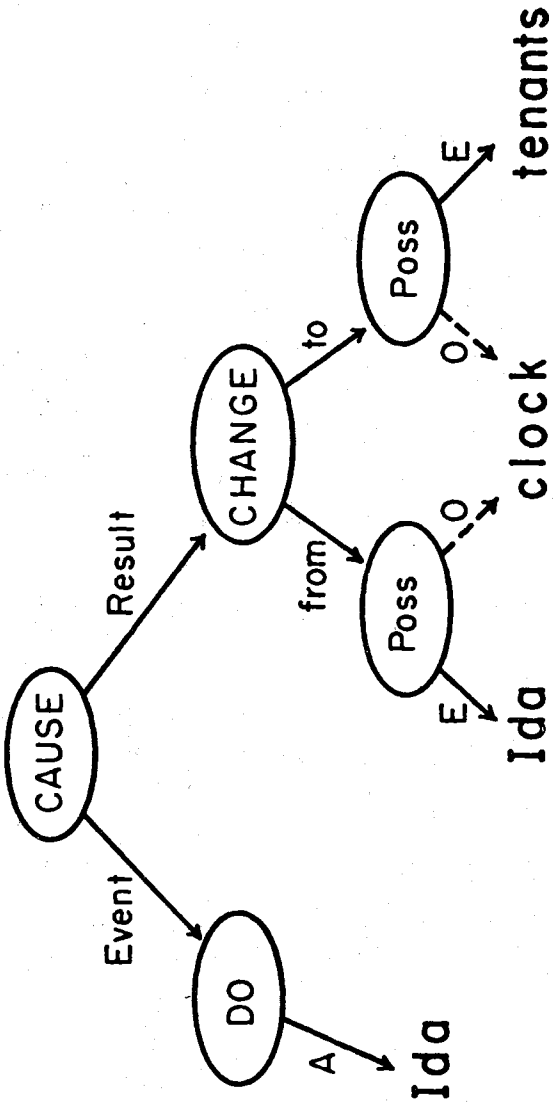
GENERAL VERB (FEW CONNECTING PATHS)

Figure 5a

Ida sold her tenants a clock
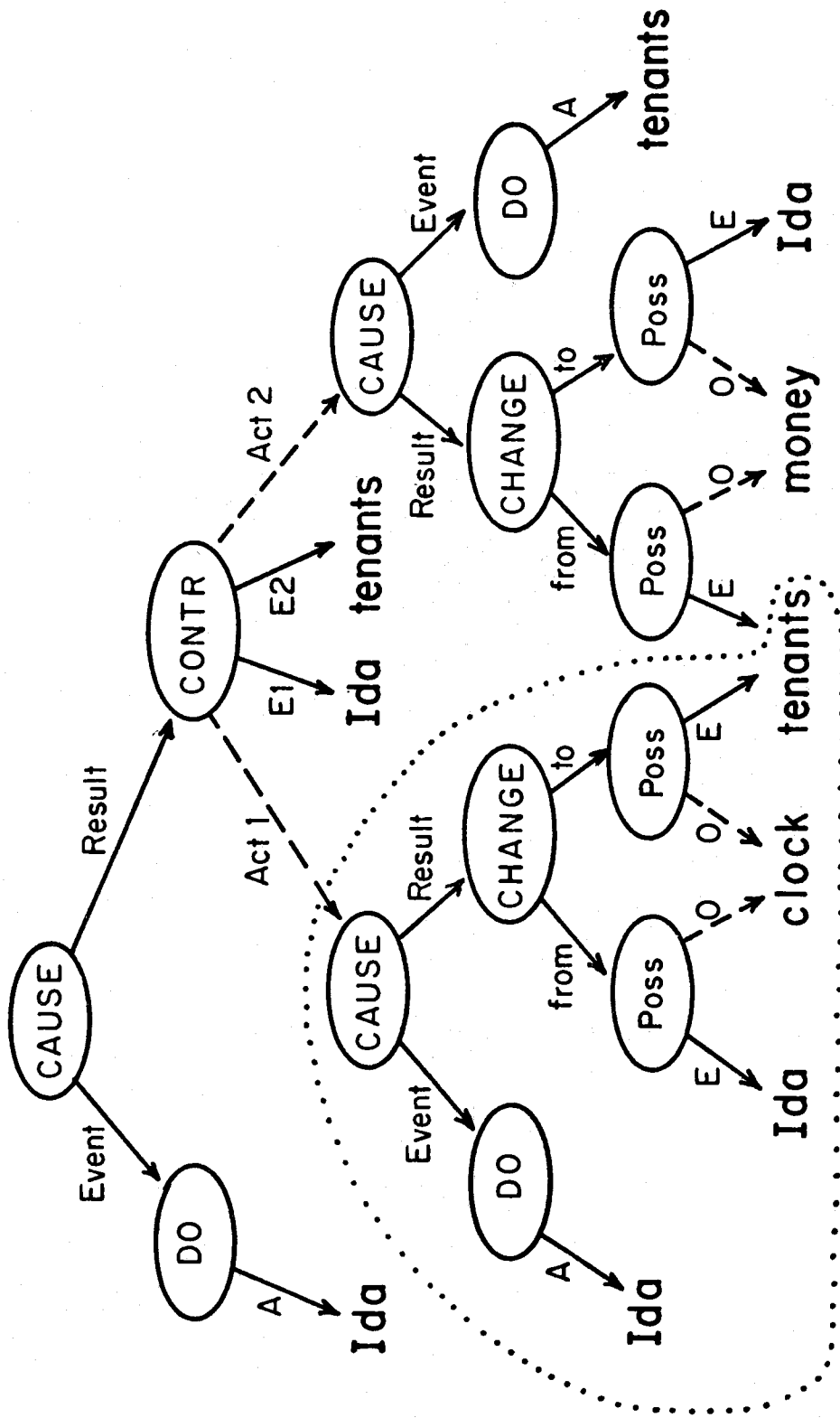
SPECIFIC VERB (MANY CONNECTING PATHS)

Figure 5b

24

25

Ida _mailed_ her tenants a clock
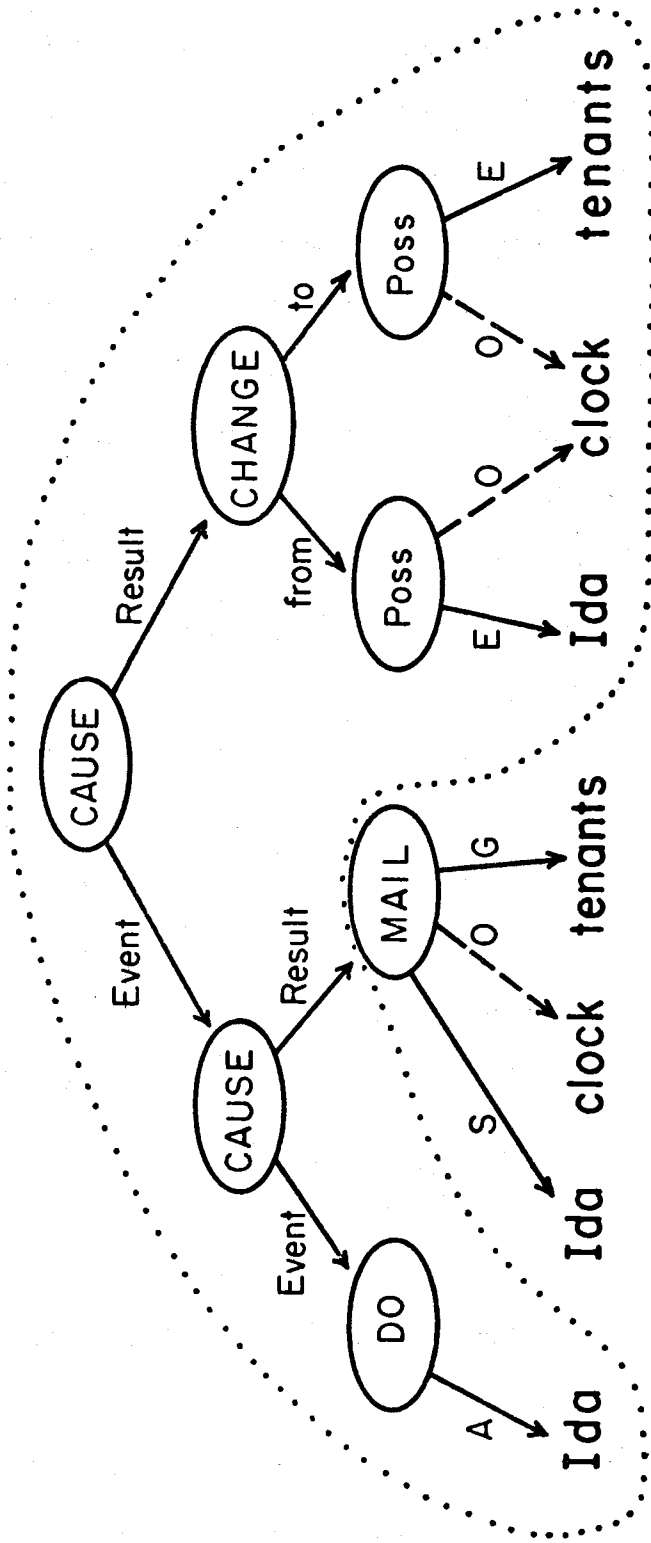
SPECIFIC VERB (FEW CONNECTING PATHS)

Figure 5c

In some complex verbs, however, the additional semantic material merely amplifies the simple verb in a manner that does not add connections between the nouns. For example, the verb mail, whose representation is shown in Figure 5c, adds the information that the method of transfer was mailing or some similar form of long-distance transfer. It is this information that makes the event description more specific than a general act of giving. However, the knowledge that the object was mailed leads to few, if any, additional connections between the agent noun (Ida) and the recipient noun (tenants). Therefore, the prediction was that use of such non-connecting specific verbs would lead to no improvement over use of general verbs in memory between the nouns.

The results were exactly as predicted: The object nouns of connecting specific verbs were recalled better than those of general verbs and non-connecting specific verbs. Thus connectivity is beneficial to sentence memory.

The experiment outlined in Table 1, in addition to testing whether complexity leads to difficulty, can be used here to measure the effects of connectivity. By varying not only the complexity of the verbs but also their connectivity, we can ask whether additional connectivity leads to improvement in memory for a text. It may be that connectivity has other beneficial effects as well: that the vividness and aesthetic value of a passage is in part due to the judicious use of highly connective predicates.

We may expect these effects to be fairly age-sensitive, if only because young children may lack understanding of some of the connective

components of word meaning. For example, at the age when the verb sell is

comprehended as having the same meaning as give, it seems likely that use

of sell leads to no improvement in text memory over use of give. Another

point that must be made is that the effects of connectivity arise from

specific representational structures. Which nouns will be held together

by a given verb must be predicted from the representational structure of

the sentence (or, for example, in Figure 5b note that sell creates many

connections between Ida and tenants, but adds hardly any extra connections

beyond that of give between Ida and clock). This line of prediction is

thus potentially much more precise than an approach that simply attempts to

divide verbs into connective and non-connective groups, or worse, into com-

plex and simple groups, without considering the semantic function of the

complexity.

e. Semantic Integrability

Another factor likely to be important in text processing is the clarity

with which the sentence meanings fit together. If the semantic structure

set up by the verbs in the sentences are such that the connections between

them are transparent--for example, if the sentence representations share

common components--then the text should be smoother and easier to

understand.

We have only begun to investigate the role of semantic connectivity

in larger texts. It has been shown that semantic integration among under-

lying subpredicates can be systematically created (Gentner, 1975). In

this study, a general verb, such as give, was presented in a passage that

contained additional semantic information, such as the fact that the giver

actually owed the money he was giving. The integration of the semantic
components of the context with those of the verb was hypothesized to pro-
duce a more complex structure--in this case, the structure of pay. As pre-
dicted, subjects hearing the extra material falsely recalled the verb which
best fit the composite structure, rather than the verb actually presented.

5. Application to a Sample Text

With the foregoing discussion in mind, let us re-examine the passage
from BABAR cited at the beginning of this section:

a. The Babar family is going to Paris.

b. Now the Babar family is on the train.

c. The train is coming into Paris.

d. Now they are off the train with all their bags.

e. The Babar family is waiting for a taxi.

f. The taxi takes them to their hotel.

g. Celeste and the children walk inside.

h. In their hotel room . . .

Figures 6a through 6g show a sentence-by-sentence analysis of the repre-
sentational structures set up by the locational verbs in the passage. (We
consider only the locational information, although the discussion could be
extended to other inferences in the passage.)
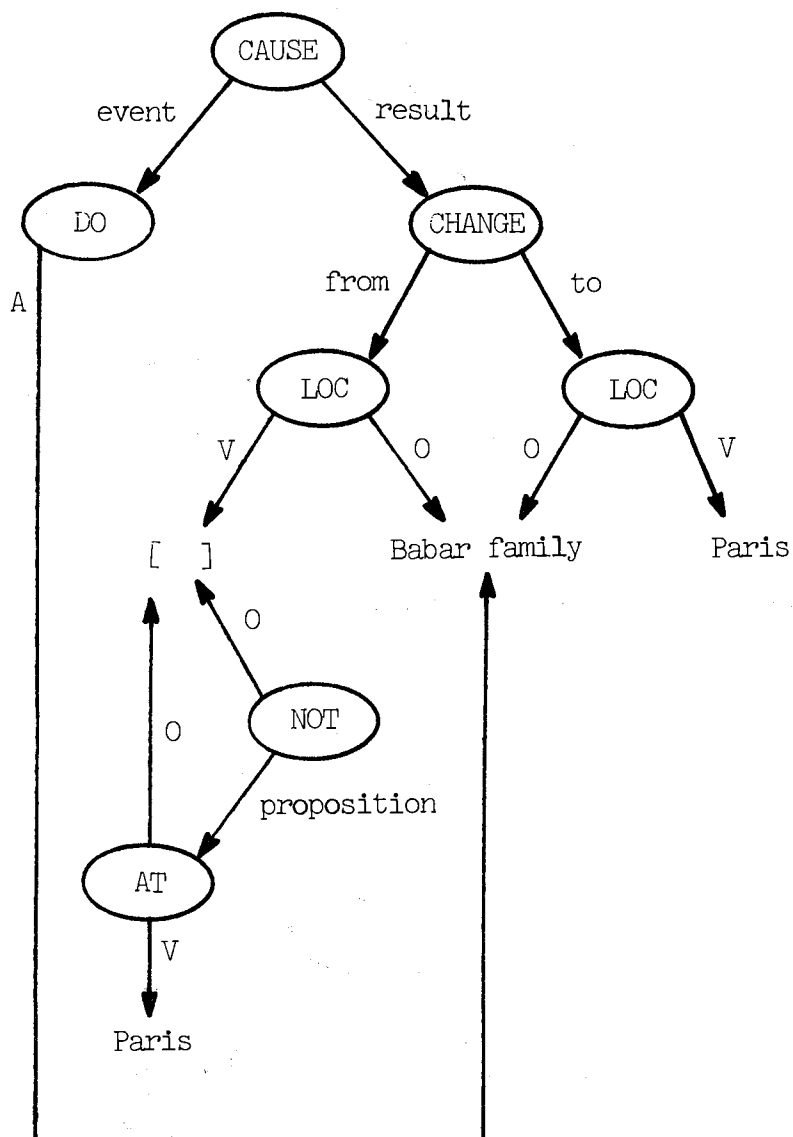
FIGURE 6a.

The Babar family is going to Paris.

Time information is not shown, since all information in the text is given in the present tense. For simplicity of presentation, in the remaining Figures (6b-6g) the information that the initial location is not the same as the final location will be abbreviated by a [?].
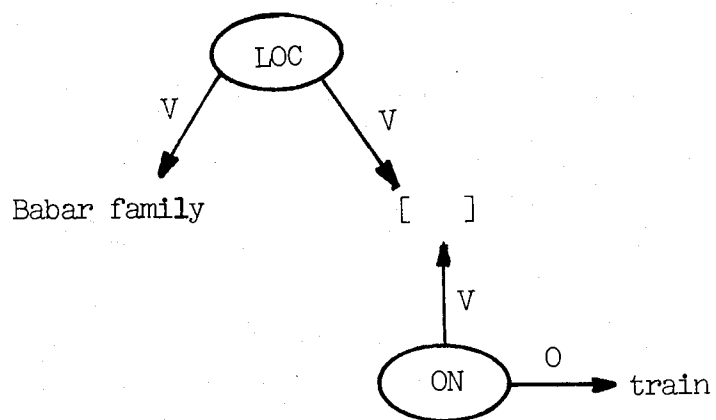
FIGURE 6b.
Now the Babar family is on the train.
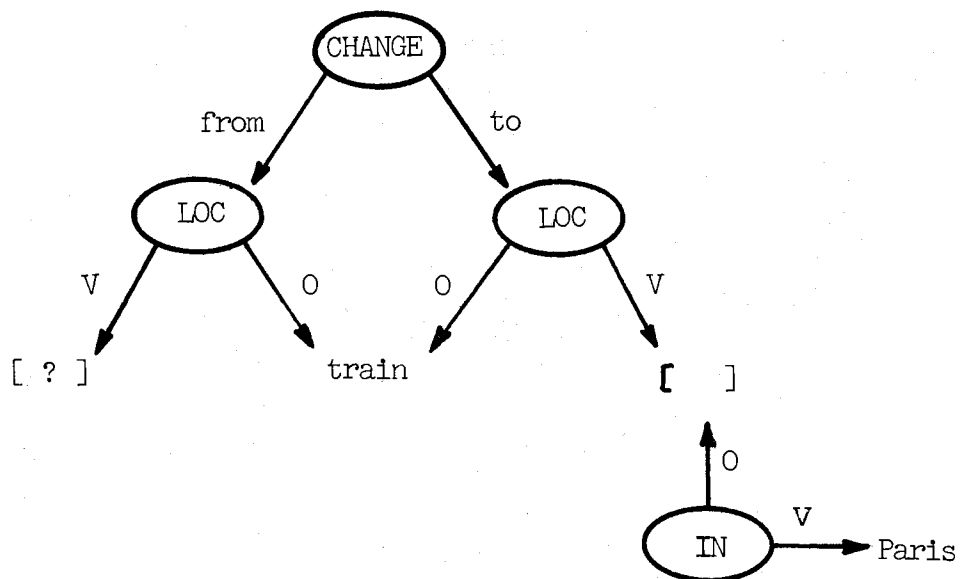
FIGURE 6c.
The train is coming into Paris.
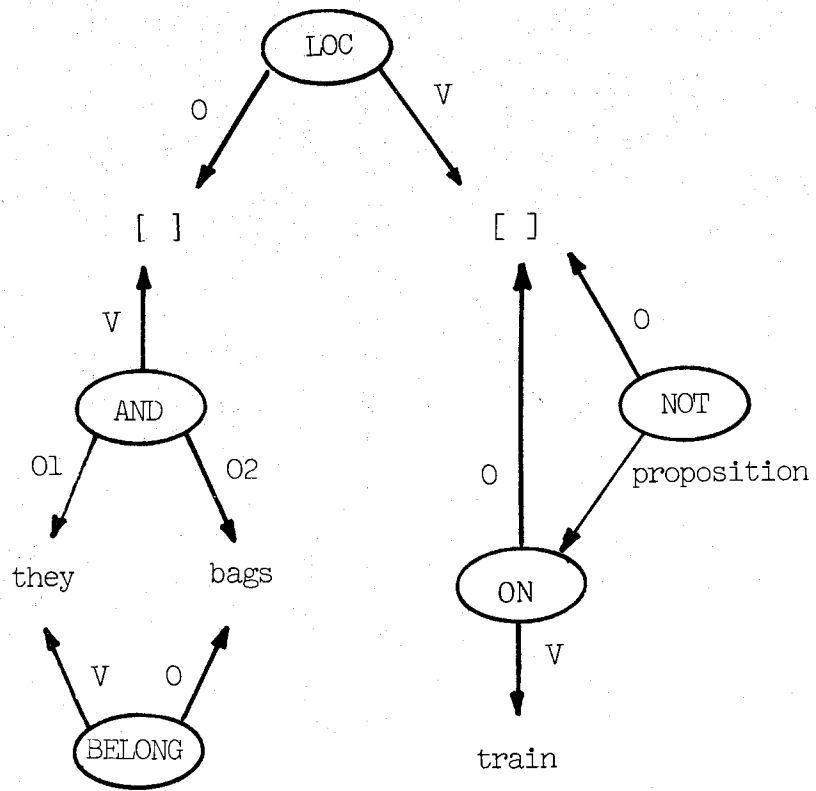
FIGURE 6d.
Now they are off the train with all their bags.
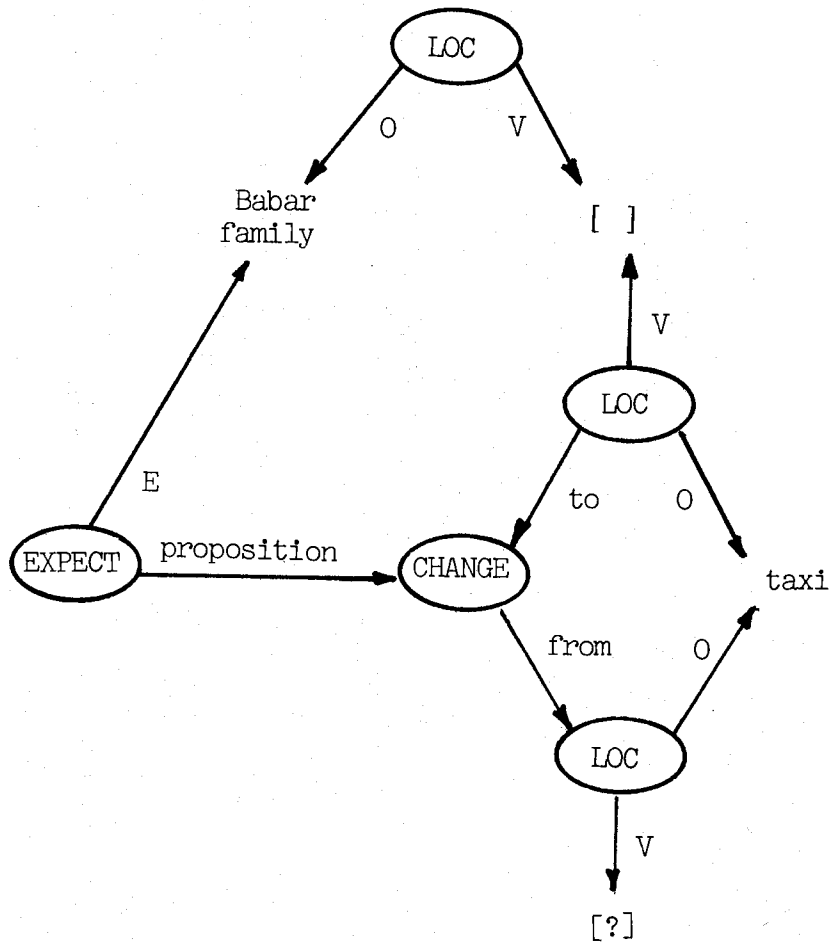
FIGURE 6e.
The Babar family is waiting for a taxi.
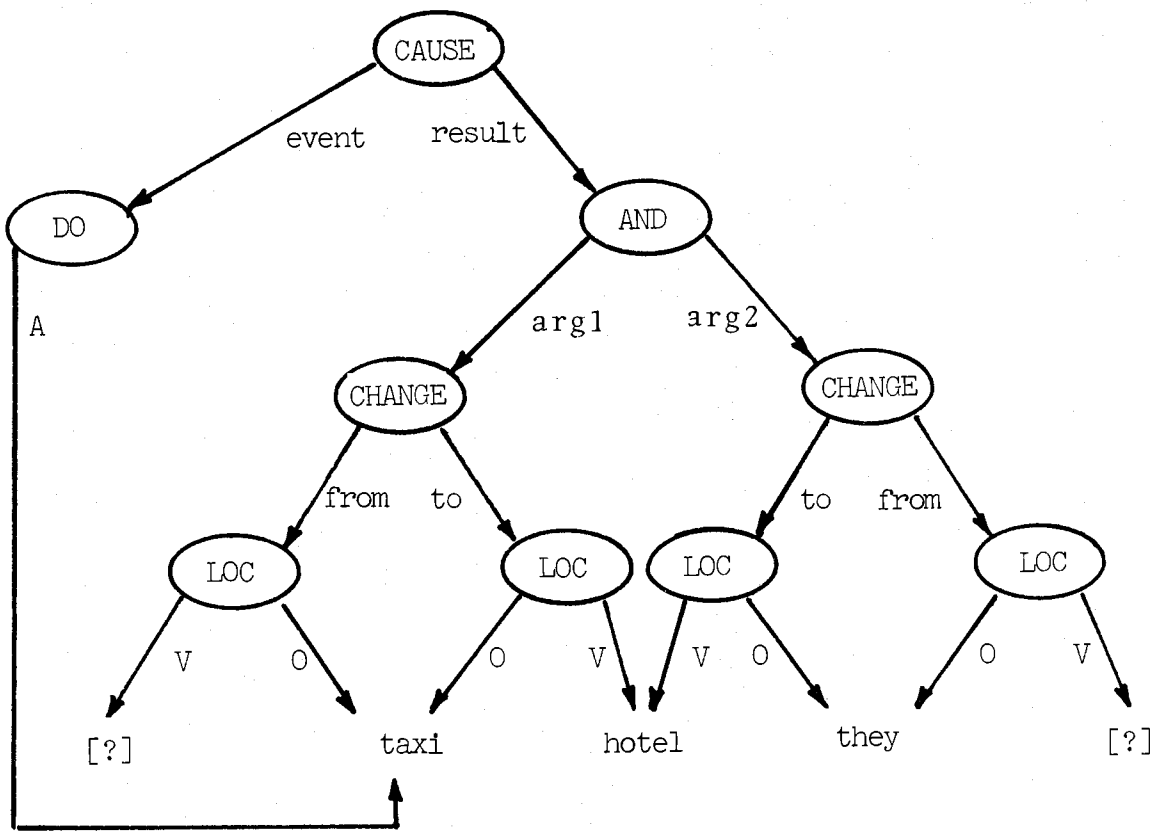
FIGURE 6f.
The taxi takes them to their hotel.
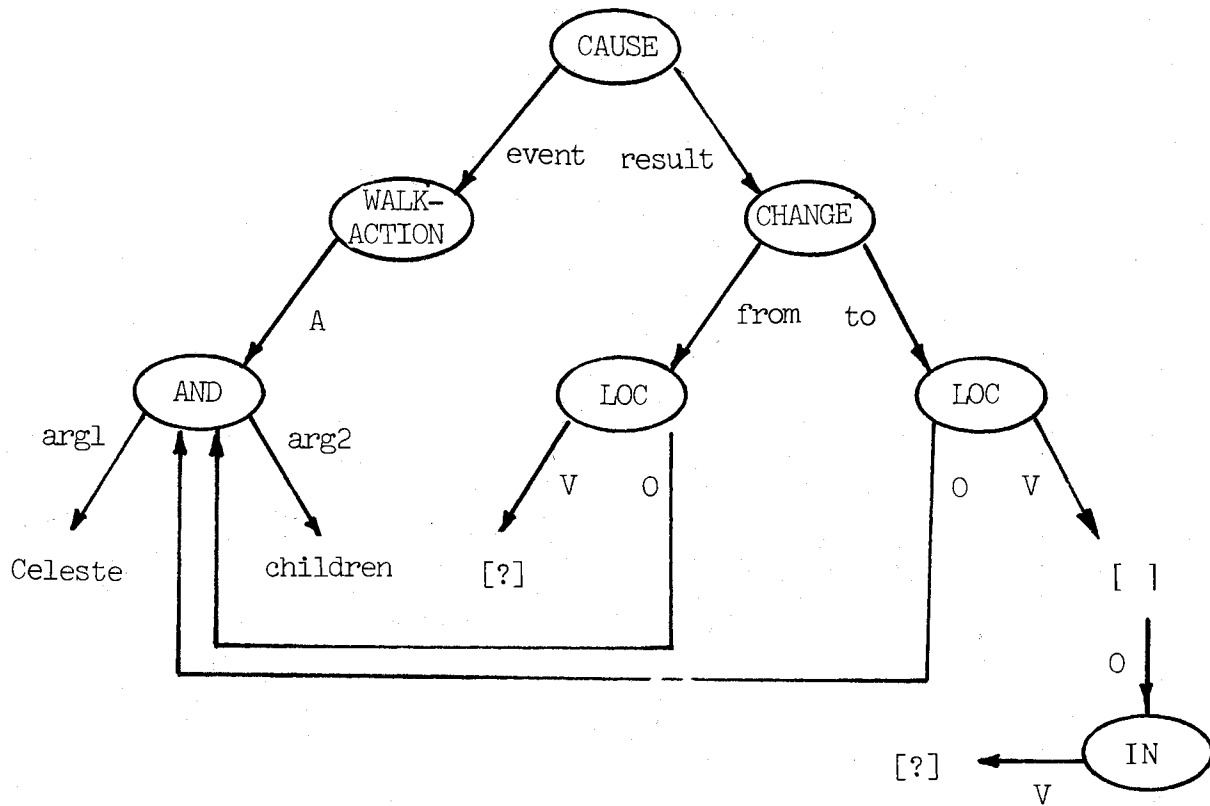
FIGURE 6g.
Celeste and the children walk inside.

Figures 7a and 7b show how these individual sentences can be combined at the intersentence level. This, of course, is the crucial process that allows readers to make sense out of passages of prose.
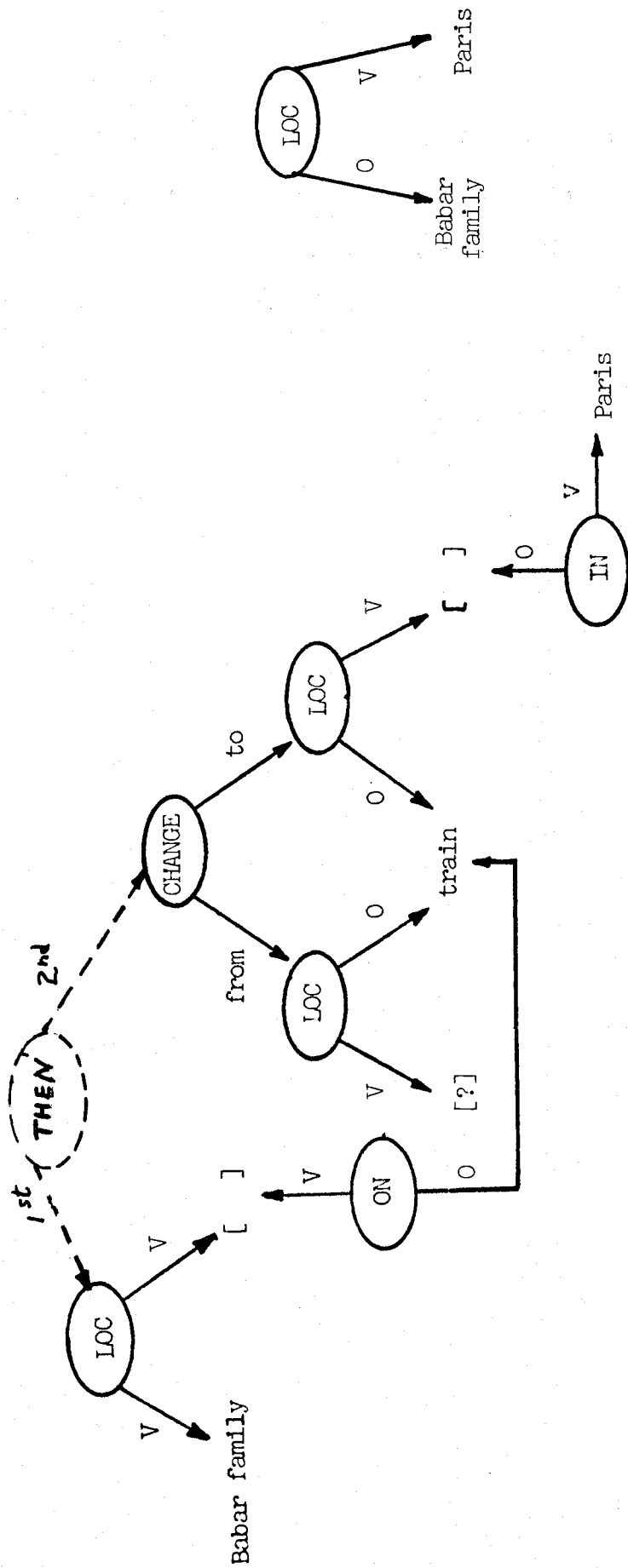
FIGURE 7a.
Tracking the Babar Family

Sentences b and c give the implication on the right.
Note: Dotted lines in the lefthand figure indicate intersentence connections.

36

Figure 7b.
Tracking the Babar Family (continued)

Sentences b and c combined with sentences d e and f.

At this point the implication can be made that the Babar

family is at their hotel in Paris, fulfilling the travel plan

in Figure 6a.
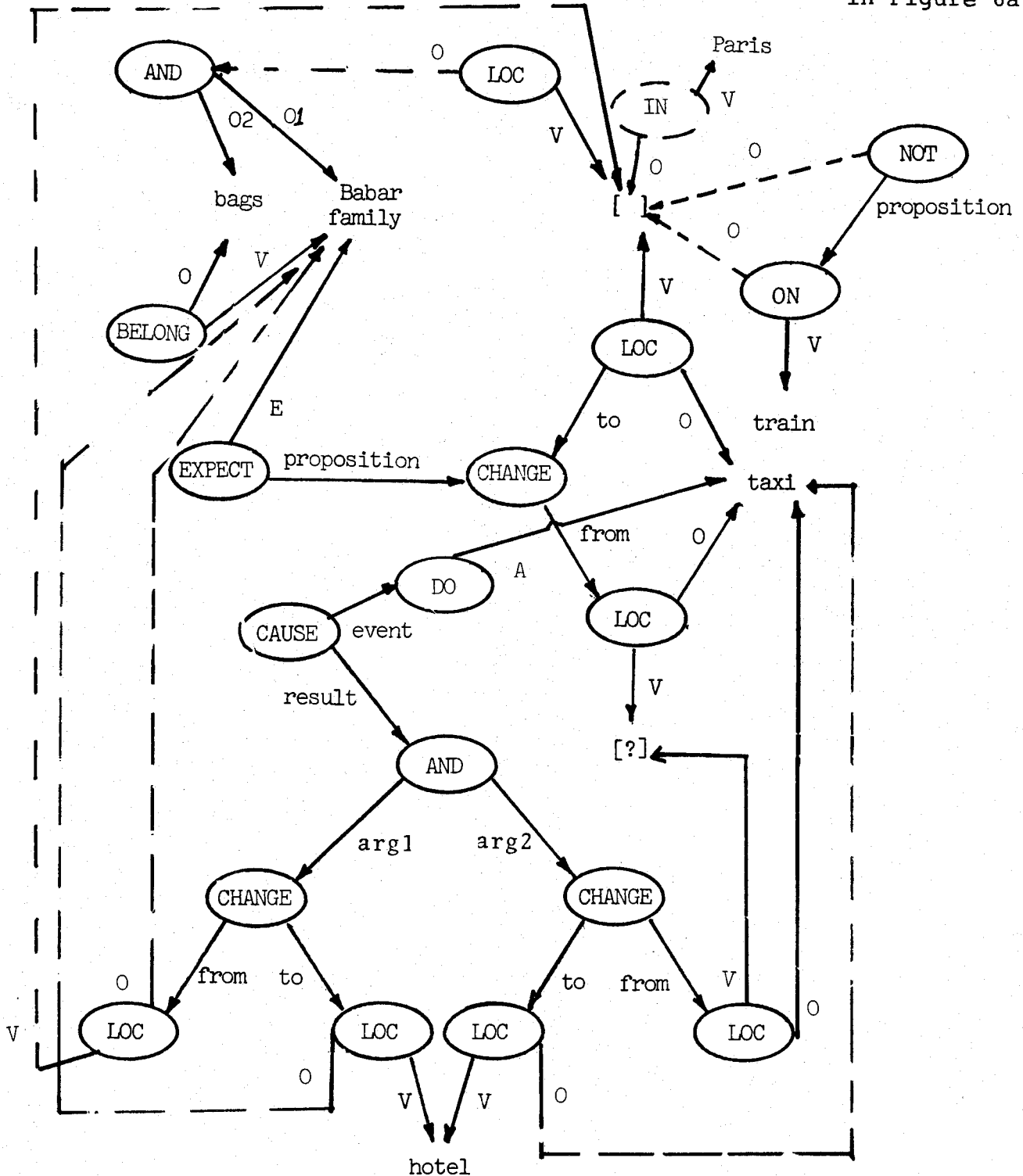
An example of incomprehensibility: The Indian Occupation. The Indian
Occupation is a story written to help poor readers obtain practice with words
that have certain patterns of spelling. It is a classic example of the kind
of confusion that results when a writer considers only the lower levels (in
this case, the orthographic level) and ignores the higher levels of text
structure. To give the reader some feeling for the problem, we reproduce the
entire passage:

> The Indians had not heard from the government. The suit for Alcatraz
> was still not settled. The Indians were discouraged and angry. They
> did not know if their goal could be reached. Some people wanted to
> tear down the buildings. "The white man is our foe," they said. "He
> took our land 300 years ago. It's true! The white man wrote treaties,
> but they were all a hoax." Other Indians said, "Wait! We must build
> a place here that we can boast about. We must have a school. It's
> dangerous for our children to roam through these old buildings. We
> need food too. We must hoe the soil and plan tomatoes, potatoes, and
> fruit." Suddenly someone roared, "Fire! Fire!" A fire had started in
> an old building. Unfortunately, the boards made good fuel. The flames
> soared high. There was no water to soak the buildings. The only water
> on Alcatraz was the drinking water brought by the boats. The Indians
> had no pumps to bring water out of the bay. Finally, the roaring fire
> was reduced to coals and burned itself out. There were no clues to
> tell how it started. Other problems came up. Food and water did not
> come when they were due. Boats cruised by, but they didn't stop at the
> island. Some of the Indians began to loaf. They forgot about their
> oath to work together. Richard Oakes decided to leave the island.
> Others said, "If he goes, we'll go too." Nobody could coax them to stay.
> Now the Indians had other foes. Time was their enemy. They had to hoard
> food and water. And there was no power or coal. The Indian people
> needed warm coats. No one would loan them money. The school had to be
> closed. Their nurse left the island. And the Indians still had not
> heard if Alcatraz could be theirs. Some of the Indians forgot their true
> goal. They argued. They fought over many issues. They roamed the
> island wondering what to do. One day a boat cruised up to move them off
> the island. Their dream had failed.

The passage seems very difficult; one finds oneself rereading
sentences and skipping backwards to find out whether one has missed prior
information. Why? It is not only because of references which come out
of nowhere, such as: Richard Oakes decided to leave the island. Another
reason for the difficulty may be the lack of overlap between the semantic

structures in adjacent sentences. For example, consider the passage that
begins

       Other problems came up.

The next sentence is

       Food and water did not come when they were due.

The verb come refers to a change of location of food and water to the
island. However, the next sentence

       Boats cruised by, but they didn't stop at the island.

refers to a distinct change of location: the paths of boats past the island.
On second reading, the connection becomes clear: it would have been desir-
able for the boats to have stopped at the island, because they might have
been carrying food and water. But none of this is transparent, and none of
it is accessible without considerable knowledge of the world. The next sen-
tence is

       Some of the Indians began to loaf.

This has even less connection with the text than the previous sentences. A
canny reader, assuming that there must be a connection, can reason out that
the Indians are loafing because they are discouraged, because there are many
problems, and so on; but again there is no semantic overlap to help the
reader with these inferences, and so again comprehension of the story depends
on considerable reading expertise.

6. <u>Concluding Remarks</u>

    Semantic structure is only one level of text structure; for a story to
be comprehensible requires much more than mere sentence-to-sentence connec-
tions. Indeed, text with good intersentence connections but poor overall
structure sounds glib and shallow, like an insincere political speech. Yet,

the semantic structures of sentences are the building blocks that make up
higher-level structures, and from which higher-level structures are de-
rived. Semantic connectivity is clearly an important factor in story com-
prehension. Further research will help to define its role more precisely,
to suggest the kinds of connectivity that are most helpful, and to make
clear the connections between sentence structures and higher-level
structures.

III. Syntactic Analysis

A. Introduction

Syntactic analysis is taken to include the enumeration, display, and
discussion of syntactic properties of texts, or more strictly, of the sen-
tences constituting texts. This entails, at the very least, a parsing (at
least one level) of each of the sentences in a text and a transformational
history of each. The rationale for parsing is discussed in section B,
along with a description of the information provided by the parsing.
Section C describes the transformational history analysis and its rele-
vance to reading comprehension.

An important consideration of syntactic analysis is a measure of syn-
tactic complexity. The difficulties inherent in arriving at a satisfactory
measure of syntactic complexity are discussed in section D.

A further syntactic characteristic of texts that may affect comprehen-
sion is the extent to which the diction employed in the text differs from
the conversational usage with which a child may be expected to be familiar.
Section E treats the analysis of differences between oral and written
language.

B. Parsing

Texts are parsed into major syntactic constituents, relationally de-
fined (e.g. in terms like subject, verb, direct object, and indirect object,
rather than purely syntactic terms like NP and VP). The following informa-
tion is included in the parsings:

a. Position of line breaks with respect to the major consti-
tuents. This is of relevance to reading comprehension insofar
as line breaks might hinder the syntactic processing of sen-
tences they interrupt. One might speculate that they would
slow processing less when they come between major constituents,
such as between the subject and the predicate, and more when
they come between a determiner and its noun in the object of
a preposition. On the other hand, if it is easier to antici-
pate what will follow a determiner than what will follow a
subject noun phrase, then line breaks that interrupt fine de-
tails of structure might require the same amount of processing
time, or the breaks between major constituents might slow
processing more.

b. Indication of whether the sentence deviates from the sub-
ject-initial order which is normal for English sentences. If
this "deviant" order is normal for that syntactic type of sen-
tence (e.g. in questions), this is also indicated. This in-
formation is indicated on the assumption that such sentences
might affect reading comprehension by disrupting syntactic
processing. This is most likely to be true for the rather
small subset of such sentences which excludes questions, quo-
tation preposings, and there- insertion sentences, as well
as adverb preposings which are not accompanied by subject-
verb inversion, since such adverb preposings often contri-
bute to the connectedness of the text by putting in sentence-
initial position a connective which relates the sentence
following it to text which preceded it. (This may be seen
fairly clearly in the parsing of BABAR.) However, in prac-
tice it turns out to be difficult in many cases to determine
whether a "deviant" order should count as "normal" for the
context or not. This is discussed in more detail in sec-
tion D (Syntactic Complexity of Texts).

c. Number of clauses or clause-remnants per sentence. This
is of relevance insofar as it may provide a measure of syntac-
tic complexity. See section D for further discussion.

d. Indication of whether the sentence being analyzed is a di-
rect quotation. This applies only to narrative texts. It is

relevant information in that one might want to be able
to segment the text into the characters' sentences and
those of the narrator, in order to analyze them separately.
Quoted speech may differ syntactically from connecting
narrative, and may require different kinds of inferences,
or more inferences, or both. It certainly should reflect
oral language more and written language less than the non-
quote portions. See Section E (Oral-Written Language
Differences) and Hermon (1979) for further discussion.

e. Comments describing 1) the exact position of line
breaks which interrupt major constituents, 2) the inter-
nal structure of constituents interrupted by other major
constituents (only a few cases of this exist in the texts
examined), and 3) the discourse function of elements con-
tributing to "deviant" order, as described above. Such
elements are usually preposed adverbials, less often post-
posed subjects.

C. Transformational History

The parsings are followed by a sentence-by-sentence analysis of the

transformations involved in the derivation of each of the sentences in

the texts. Such an analysis necessarily presupposes a particular theory

of transformational grammar. In compiling the lists of transformations

we have tried to include only transformations which the most generally

accepted versions of transformational grammar would include. That is, we

have listed only relatively uncontroversial transformations which affect

major syntactic constituents, and have avoided discussion of such de-

tailed and relatively uncharted areas as the derivation of possessive

constructions and pre-determiners.

This information is relevant to reading comprehension insofar as trans-

formations contribute to difficulty or ease of syntactic processing. There

are two ways in which they might do this. First, it could conceivably be

the case that certain transformations contribute to the generation of structures which are unfamiliar to the reader, or which for other reasons present processing difficulties. We have in mind here such transformations as Relative Clause Extraposition, which moves a relative clause from the noun phrase it modifies to the end of the clause that the noun phrase occurs in. This particular rule is likely to present processing difficulties for the young reader for both of the reasons mentioned above; the resultant construction is one which the young reader may never have heard, and it separates syntactic units which are very closely bound semantically. Reuniting them may be a relatively difficult task for a young child. If we had firm knowledge of which transformations were responsible for structures which would be likely to cause processing difficulties, then an index of the transformations involved in the generation of a text could provide the data base for computing a measure of difficulty for the text. Such information could, it would seem, be obtainable from a carefully designed series of experiments.

D. Syntactic Complexity of Texts

Several methods of calculating syntactic complexity were employed in the first analysis (BABAR and DESERT). None was satisfactory. The first were measures which refer to the number of clauses or verb-containing clause remnants per sentence, with sentence defined as material beginning with a capital letter and ending with a full stop (period, question mark, or exclamation point). Clause, as used here, refers not only to independent clauses, subordinate complement clauses, and relative clauses, but also to any clause remnant which contains a finite, infinitive, or participial verb

form whose objects have the same form and position relative to the verb

they would have if their verb was a main verb. Thus, in BABAR there are

sentences which have no clauses, since there are fragments consisting sim-

ply of NPs which are punctuated like sentences, e.g. 115 and 155:

      115.   Poor Babar!

      155.   The Mustache-man!

Verb-containing remnants include infinitive clauses/phrases which function

as objects, adverbials, noun modifiers, etc., and participial modifiers and

complements. Nominalizations and preposed adjectives are not included, on

the grounds that the object in such constructions does not have the same

form or position that it would have if its verb were a finite form.

    Two statistics were calculated: range of verbs (defined as above) per

sentence and average number of verbs per sentence. On the first measure,

BABAR and DESERT are nearly equally complex, since the range in BABAR is

0-3 and the range in DESERT is 1-3. On the second measure, DESERT is con-

siderably more complex, since the average number of verbs/sentence is 1.48,

whereas in BABAR it is only slightly more than 1--1.05 to be precise. This

corresponds to our intuitions about the complexity of these two texts. By

comparison, the Reader's Digest article from which DESERT was written, a

piece called "The Challenge of the Desert" (CHALLENGE), had a range of 1-6

verbs per sentence, and an average of 2.30.

    The measure of complexity provided by these figures is not fine enough

to be of much use. While CHALLENGE is intuitively quite complex syntactically,

probably well above the abilities of the average second-grader, books are

written for young children which do not seem particularly syntactically com-

plex, but which have similar complexity indices if complexity is mea-
sured this way.   For example, Charles M. Schulz's <u>He's Your Dog, Charlie</u>
<u>Brown</u> (1974), hereafter termed DOG, has a range of 0-6 verbs per
sentence, and an average of 2.44 verbs per sentence.   The average num-
ber of verbs per sentence is even higher than CHALLENGE, although intui-
tively DOG seems much easier.

The number of verbs per independent clause would seem to offer a
promise of being a more useful measure of complexity, since it would
not treat the number of independent clauses conjoined in a sentence as
contributing to complexity, which verbs/sentence does.   However, a com-
parison of the four texts for which the calculation was made indicates
that verbs/clause is no more useful than verbs/sentence, in that both
make DOG appear even more syntactically complex than CHALLENGE.

	Verbs/sentence	Verbs/clause
BABAR	1.25	1.31
DESERT	1.48	1.45
DOG	2.44	1.75
CHALLENGE	2.30	1.61

Clauses per T-unit and T-units per sentence (Hunt, 1965) were
also calculated to see if they would provide an index of complexity
that corresponded to our pretheoretical judgments of the relative com-
plexity of these four texts.   But they turned out to be no better than
any of the other measures.   In fact, as indicated in the chart below,
where the books are listed in order of increasing complexity according

to our judgments, the figures for T-units/sentence indicate that
CHALLENGE is considerably less complex than even BABAR and DESERT, a
claim we cannot accept as accurate. (T-units/sentence is supposed to be
inversely proportional to complexity, and is supposed to reflect the
achievement of multi-clause sentences by coordination rather than sub-
ordination.) Clearly this ratio can only be sensibly interpreted in
conjunction with some measure of sentence length in terms of words or
clauses. But as it turns out, only measures of sentence length in
terms of words per T-unit or words per sentence provide indices of
complexity which rank the four texts in the same order as we did. Of
these, words/sentence appears to be truer to our judgments of the rela-
tive distance between the texts, contradicting Hunt's claim of supe-
riority for words/T-unit as a more accurate measure of maturity of
style than words/sentence. Needless to say, such measures give no in-
dication of the source of the complexity that they indicate.

	Clauses/T-unit	T-units/sentence	wds/T-unit	wds/sentence
BABAR	1.19	1.03	6.97	6.58
DESERT	1.13	1.02	8.49	8.56
DOG	1.32	1.30	11.15	12.91
CHALLENGE	1.30	1.22	13.50	16.43

(Interestingly, the clauses/T-units ratio for CHALLENGE is identical
to that found by Hunt in writing samples from fourth graders. The
twelfth graders whose writing he examined averaged 1.68 clauses/T-unit.)

Among the syntactic constructions contributing, in our opinion,
to the complexity of CHALLENGE which are overlooked by these measures

are: 1) Sentences with adverbial phrases as subject: <u>From dawn to</u>
<u>dusk in the sun can finish him</u>. 2) Sentences where the subject is in-
verted after a preposed locative adverb: <u>Yet in every desert are living</u>
<u>things--plants, animals, human beings--which defy the dictionary and</u>
<u>the desert itself</u>. <u>Today around his villa are green lawns with foun-</u>
<u>tains playing</u>. 3) Sentences with dangling pre- and post-modifiers.
(We do not claim that these are ungrammatical, or even poor style,
only that their correct interpretation is not transparent.) Some
examples: <u>Following the Nile's course for 560 miles, from near Luxor</u>
<u>to the Delta, it is six miles wide</u>. <u>A species of frog in the Austra-</u>
<u>lian desert stores water in its abdominal cavity, bloating itself</u>
<u>until it is spherical in shape</u>. <u>The world's largest, the Sahara, is</u>
<u>advancing on a 2000-mile front, in some places as much as 30 miles a</u>
<u>year</u>. <u>This fringe is invaded by nomadic herdsmen, forced out of the</u>
<u>desert</u>. 4) Sentences where the use of semicolons indicates a con-
nection between ideas, but does not make that connection explicit, as
in these passages. In the first, the second clause is an exemplifica-
tion of the claim of efficiency made in the first. But it is not clear
even to us what the connection is supposed to be in the second passage,
and there can be little doubt that the syntactic juxtaposition of
clauses without explanation (or even a hint of the motivation for juxta-
position) causes difficulty. <u>The cactus has an efficient water-storage</u>
<u>system in its thick stem; one species of tree cactus may contain hun-</u>
<u>dreds of gallons</u>. <u>Egypt once was a semi-desert; then in the seventh</u>
<u>century the invading Arabs brought in great herds of camels, sheep</u>
<u>and goats</u>.

48

Another possible measure of complexity refers to the proportion of sentences which deviate from the canonical subject-first order characteristic of English. Deviant sentences, defined this way, include questions and sentences containing quotations and preposed adverbs, regardless of whether the subject is inverted to follow the verb or not. Thirty-one percent of the sentences in BABAR and thirty-seven percent of the sentences in DESERT have non-canonical word order. But percent of sentences with non-canonical word order is a very rough measure since many such sentences are syntactically unmarked in the sense of psychological marking. For example, questions are pragmatically unmarked (in the same sense), in that the non-deviant order of constituents would make the text bizarre or incoherent.

Defining "deviant" is the major problem with this measure of complexity. The strictest criterion for "deviant" order--that an apparently deviant order of constituents in a sentence should be considered natural only if the sentence would be ungrammatical with any other order--is too strict to be useful. In addition to labelling as deviant inverted quotation-containing sentences like "Don't worry," says Celeste, it does not allow one to consider questions with the order Object-Verb-Subject, such as sentence 20 in BABAR:

20. What is this?

to be natural, given the existence of nonsynonymous grammatical echo-questions with the order Adverb-Verb-Subject like Where is he? Furthermore, it would categorize sentences which begin with a connective or deictic adverb like so, then, later, etc. as deviant if the adverb would be

grammatical in some later position, despite the fact that in such a position it would detract from the connectedness of the text.

One might suppose that the number of transformations in the derivational history of a sentence would provide an index of its syntactic complexity, but this is unlikely to be true, for the simple reason that while some transformations, such as Passive, Gapping, and Niching, may very well contribute to syntactic complexity and difficulty of processing, others, like Pronominalization, Equi-NP-deletion, and WH-movement, most likely do not have this effect. To utilize the transformational history of a sentence to provide an index of its complexity, one would need fairly certain knowledge about which transformations contribute to syntactic complexity and which ones reduce it or are neutral. We do not have such knowledge at this time; it may turn out that transformations do not have any absolute complexity coefficient. It may be, instead, that whether a certain transformation contributes to the complexity of a certain sentence will depend on the discourse function that sentence is intended to serve (cf. Green, 1978). This is a burgeoning field of research, but there are no firm results as yet.

E. <u>Oral-Written Language Differences</u>

This section treats the analysis of differences between the language of written texts and the language of speech. Such differences may affect reading comprehension in at least two distinct ways. First, during the period before decoding proficiency is complete, texts in a specifically written-language style may present what amounts to the added burden of learning a foreign dialect at the same time the learners are supposed to be mastering the relation between the language they are already familiar with and

classes of marks on paper. The langauge of written texts may be assumed
to be unfamiliar and foreign to children to the extent that they were not
read to regularly as preschoolers. Children who come across unfamiliar words
or constructions before decoding is mastered may not know whether to doubt
the familiarity of the form, or their own mastery of the sound-grapheme cor-
respondences they are supposed to be learning. We speculate that children who
have been read to extensively as preschoolers, and are at least passively
familiar with the language of written texts, will not be much bothered by
the differences between oral and written language, but that instances of
specifically written language in texts may provide a source of difficulty
for children who are relatively unfamiliar with books and written stories.

Second, after decoding proficiency is attained, too much oral language
in a text or oral language in unexpected places may slow comprehension or
reduce it in two ways. In confounding the reader's expectations of exclu-
sively written language, there may simply be a jarring effect of inter-
rupting the cognitive processes which contribute to comprehension. The
result of this, we assume, would show up largely as slower comprehension
rates. But a child might also be distracted in a more serious way--the
child may begin to look for an implicature of this inappropriate style and
become sidetracked, wondering why the text was written as if it was someone
talking.

Carefully designed experiments with readers of various proficiency
levels and backgrounds should be performed to test these hypotheses, since
they have clear and relatively easily implemented implications for the
writing and editing of texts for different types of readers. Among the

things that would have to be controlled in such experiments are the native dialect of the reader and the reader's familiarity with the language of books.

IV. Analysis of Text-Level Properties

This section contains discussion of the analysis of text-level properties of texts which may affect reading comprehension. Section A treats the analysis of the organizational structure of texts, and contains comments on how this structure is indicated overtly to the reader, and how well it is indicated. Section B is devoted to discussion of a number of text-level discourse properties. These include information structures, rhetorical structures, and connective devices, including anaphoric (pronominal) devices.

A. Text Structure

1. Introduction

The structural analysis of texts can be carried out at a number of levels, depending on the length and type of text. A richly structured text such as BABAR could possibly bear analysis at as many as four or five levels of organization. In the absence of a general and empirically vulnerable theory of the structure of texts (cf. Morgan & Sellner, in press, for discussion), it has been necessary to turn to pre-theoretical analyses and analyses restricted to certain literary genres like the Plans analysis described in Section IV. A. 2., and the Story Grammar, discussed in Section IV. A. 3., which are both limited to analysis of narrative texts, the latter to narrative fiction. Absence of theory may be less of a problem for expository texts, since the logic of the argument should largely

determine the structure of the text, and a well written text will make this
structure explicit, both at the sentence level, and at higher levels of
structure. Where a large number of inferences are required to connect in-
dividual sentences and paragraphs with each other, and where there is an
absence not only of explicit connectives, but of syntactic or lexical clues
with which intended conceptual relations between parts could be inferred,
we would predict that comprehension will be very poor comparatively.

2. Interacting Plans Analysis

a. Introduction

If we want to know why some children have difficulties learning to
read, we must develop a better idea of what they are reading, or could be
reading. Analysis of texts alone will not tell us what is the best way to
help someone learn to read, but it makes asking better questions possible.
This section, together with the reports, "What makes a good story?" (Bruce,
1978) and "Interacting Plans" (Bruce & Newman, 1978), constitute one part
of the larger text analysis effort.

Understanding the plans and beliefs of characters in a narrative is
clearly an important aspect of reading comprehension. The study of such
plans is part of text analysis because it depends on an exacting study of
the way actions are described in the text. It is also an analysis of what
goes on "between the lines" of a text since it considers motivations and
reasonings that may not be explicitly stated.

A plans analysis is particularly appropriate for two of the central
questions of text analysis: "What is it that makes a given story easy or
hard to comprehend?" and "What is it that makes a given story good or bad?".

The answers to these questions surely interact, but it is not true that difficulty alone determines the quality of a text, nor even its appropriateness for a child whose reading skills have developed to a known level. The concluding section deals with this issue further.

b. An Interacting Plans Analysis

An important aspect of a narrative text is that it relates actions connected through goals, effects, and enabling conditions. The statement, "We understand actions in terms of goals," has become a truism; actions simply are the way goals are attained. This is true for the realms of conversation, stories, or human activity in general, and there has been extensive work to show just how goals and actions relate. But an important implication of goal-based understanding of actions is often overlooked. If we can interpret an action in terms of goals, then so can others who may be affected by that action. They may then act, not just in terms of their goals, but in terms of their understandings of the actor's goals. This means that when two or more people inter-act, their plans can reach a level of complexity that is difficult to foresee from consideration of single actor plans.

The distinction between simple plans and interacting plans can be seen in the first part of the Grimms' fairy tale, "Hansel and Gretel" (Appendix 1). Hansel and Gretel are the young children of a woodcutter. Their step-mother convinces the woodcutter that they have too little food for both the parents and the children. Her solution to this problem is to take the children into the woods and abandon them. The execution of her plan starts the first episode of the story, an episode that is analyzed in detail.

A simple analysis of "Hansel and Gretel" would show that Hansel has a goal--to be able to return home after being taken into the woods. To reach that goal Hansel drops pebbles along the trail so that the children can retrace their steps. We could understand what Hansel does in terms of a plan in which dropping pebbles is an action appropriate to the goal. The plan would show how the actions of dropping pebbles and following the trail fit together, and how they produce desired outcomes for Hansel and Gretel.

But such a plan would be incomplete. Hansel and Gretel are being taken into the woods deliberately by their parents. Hansel knows that he should drop pebbles because he and Gretel have overheard their parents plotting against them. The children's plan is a response to their conception of their parents' plan. Hansel and Gretel are not just "returning home" but are "countering" the plan they perceive their parents to have. It would have been of little use for Hansel to drop pebbles on a familiar trail, and if his parents were planning to kill the children outright, some other response would have been more effective. His action becomes meaningful only with respect to his perception of the structure of his parents' plan.

What we find upon further analysis is that each of the characters in the fairy tale is acting in a reality determined by his or her perception of the others' plans. They continually evaluate what the others are doing and react accordingly. Such behavior, characterized by interacting plans, is fundamentally different from that found for one-person plans. The discussion to follow develops this idea further. It relies on ideas developed in work on planning algorithms (e.g., Tate, 1975; Sacerdoti, 1973, 1975; Sussman, 1975), on plan recognition (e.g., Schmidt & Sridharan, 1976), and on the use of plans (e.g., Cohen & Perrault, 1976; Bruce, 1979), but the

focus is on the interactions among plans. The problem is not just to show

how actions can be organized into plans. We need to do that, but we also

need to consider concepts such as cooperation and conflict, actions defined

in terms of plans, intentionality, degree of interaction, levels of concep-

tualization, and beliefs about plans. The next section presents a simple

example of an interacting plan, taken from the analysis done in Bruce and

Newman (1978).

c. A Simple Interacting Plan: A "Request"

> When they reached the middle of the forest, their father
> said, "Now, children, pick up some wood. I want to make
> a fire to warm you."
> Hansel and Gretel gathered the twigs together and soon
> made a huge pile. Then the pile was lighted, and when it
> blazed up the woman said, "Now lie down by the fire and
> rest yourselves while we go and cut wood. When we have
> finished we will come back to fetch you."

The parents' plan is an interacting plan, since it is a plan to achieve

goals in interaction with the children. The idea of an interacting plan can

be illustrated with a simple example (see above) taken from their overall

plan. Figures 8 and 9 represent the parents' plan to build a fire for the

children once they are out in the forest. Figure 8 represents the parents'

plan to keep the children warm while they are waiting for the parents to

finish cutting wood. With this example we can illustrate some of the basic

nodes and relations used in the representation system. Then, using Figure 8

in combination with Figure 9, we can illustrate a simple example of inter-

acting plans.

One action ("Parents light pile of twigs") is shown in Figure 8 in the

square node. Fire burning is a simple state which satisfies the intentional

state (labelled "P.A") which is the mental state leading directly to the act.

The IntendAchieve state is specified by an IntendMaintain state which in this case is the more general intention to keep the children warm. Since the parents know that a burning fire will Produce warm children, they know that the general goal of keeping the children warm can be accomplished in this case by causing a fire to be burning.

57
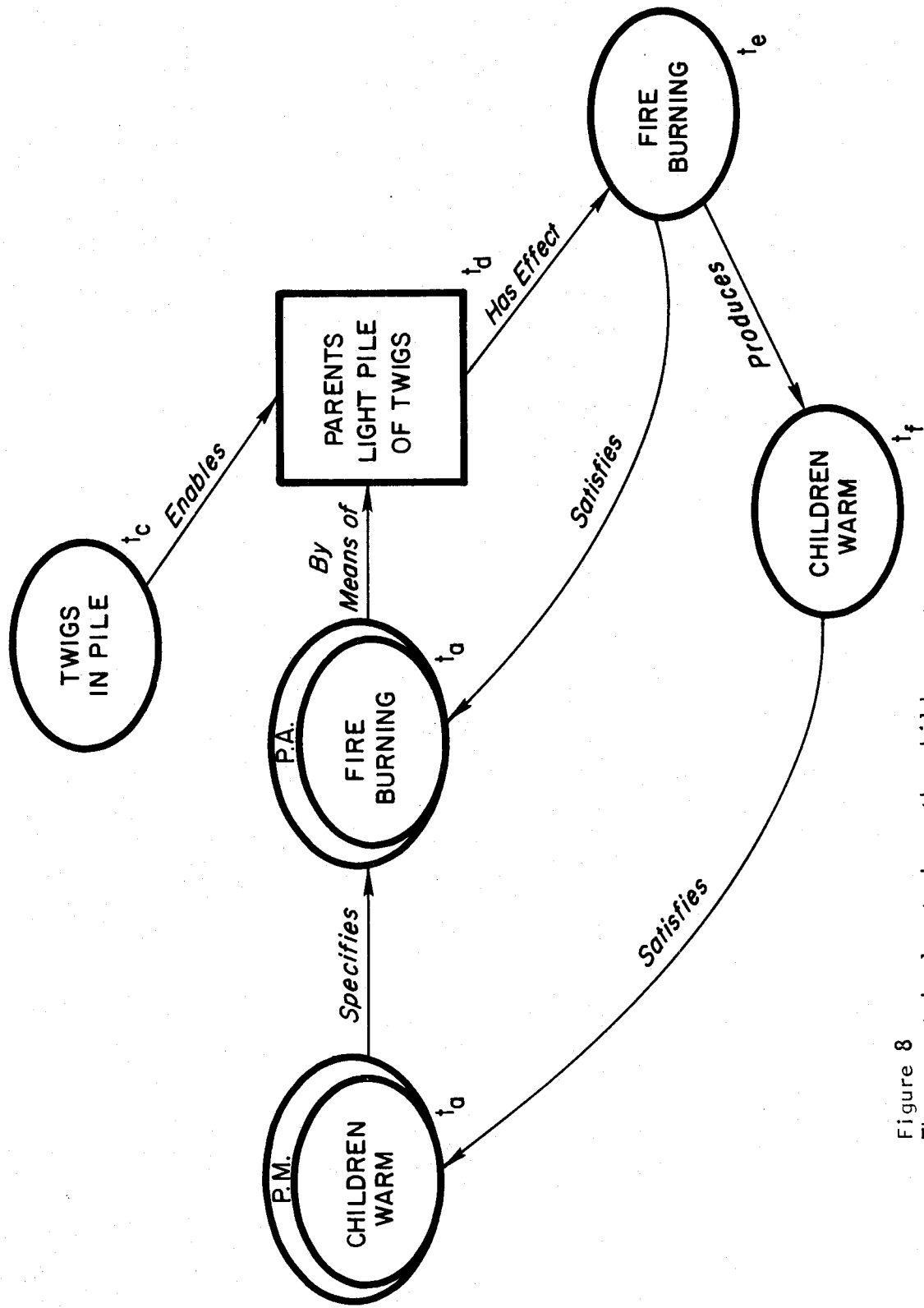
Figure 8
The parents' plan to keep the children warm

The link to Figure 9 is the state "Twigs in pile." This is a neces-
sary condition for the pile of twigs to be lit so it is linked to the act
of lighting by an Enables relation. Whenever an act has an enabling condition
that is not met, an intention to achieve that state is generated. In Figure
2 that intention is represented as an intention by the parents to achieve
the state of a pile of twigs existing. In another situation this goal might
be achieved by going about gathering twigs. But here, the parents choose to
get the children to perform the necessary actions. Thus, we have an elementary
interacting plan. The intention to achieve a pile of twigs is changed into
an intention to achieve an intentional state in the children. This new goal
is achieved by means of saying to the children, "Now, children, pick up some
wood. I want to make a fire to warm you." This, of course, is a request, and
it has the effect of the children having the intention to achieve a pile of
twigs by means of gathering twigs. This action satisfies the parents' inten-
tion to have a pile of twigs and satisfies the enabling condition for their
building a fire. Notice that while the parents' intention to have twigs in a
pile is present at time a, the children's intention comes into existence at
time b, only after the parents' request.
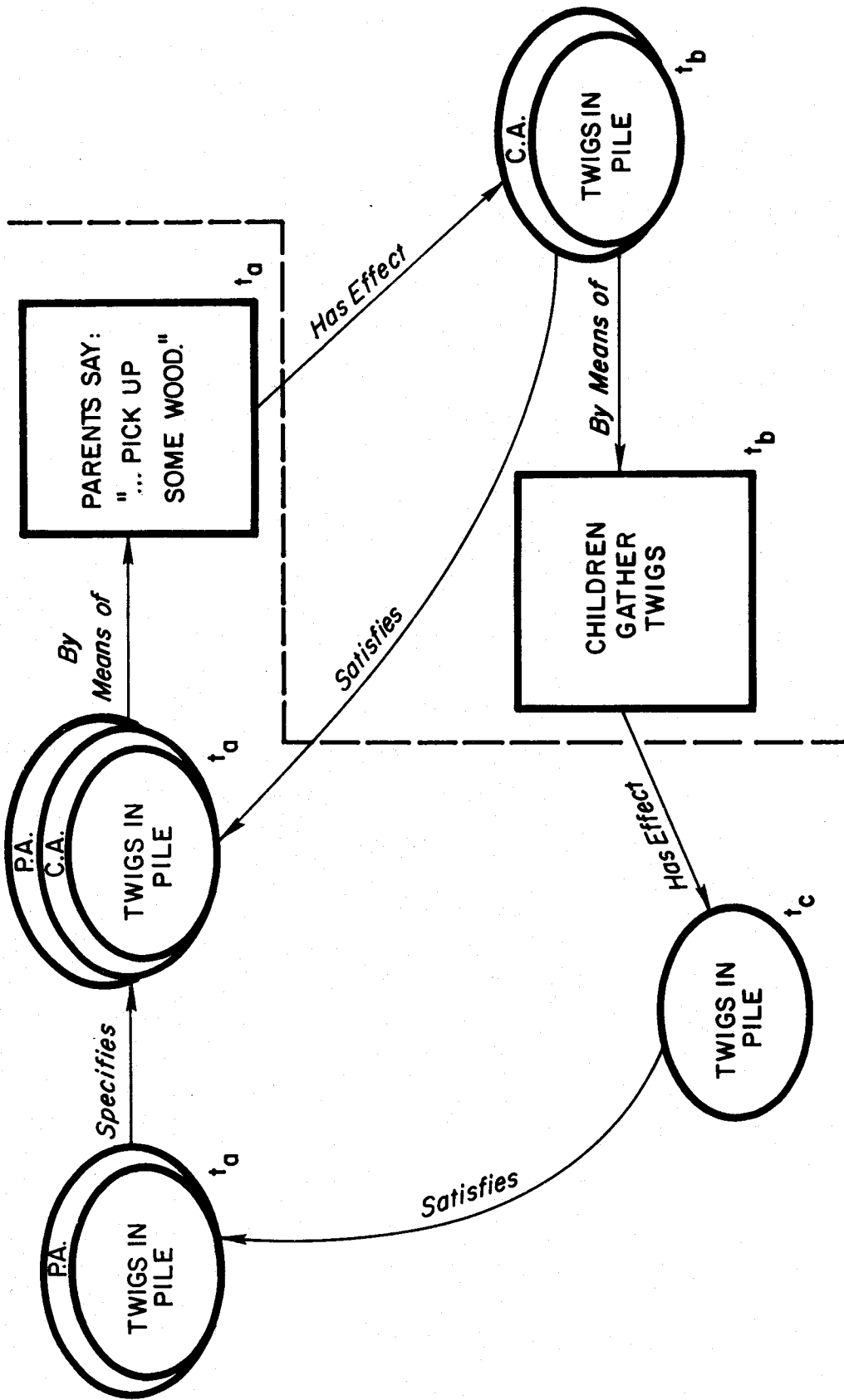
Figure 9

A simple interacting plan: Requesting help from the children

d. Virtual Plans

    One of the most striking things that emerges upon even a cursory exam-
ination of simple stories or conversations is that people frequently act
in the context of complex plans. Even more striking is that they often act in
response to or with respect to plans that they perceive being carried out, or
that they intend for someone else to carry out, that is, plans which have the
structure of actual plans but are not in fact used as determinants of behavior.
For example, in one segment of natural conversation (collected by William Hall),
an analysis in terms of plans shows that a mother in a mother-child conversa-
tion is acting in response to her perception of the child's plan and the plan
she wishes the child to have. Either or both of these plans may differ from
the plan the child actually has.

    The result is a much more complex structure in which plans themselves be-
come units. Consider, for example, the story of "The Dog and the Cock" (Ap-
pendix II). In it, a fox tries to entice a rooster to come down from a tree
by flattering him and inviting him to "come to breakfast." The rooster accepts
the invitation with the proviso that he may bring along his friend, who happens
to be a big dog. The fox greedily agrees but soon finds that the dog is not
another rooster (= breakfast) but a foe who bites him on the nose. Clearly
there is a plan of the fox to eat the rooster. At the same time, we have the
rooster's plan to get rid of the fox. These plans interact with each other in
terms of a third plan which neither of the characters expects to carry out nor
intends to have carried out.

    The virtual plan is that the rooster comes to the fox's house as a guest
for breakfast. The rooster, we may assume (though many children in fact do

not assume this), does not intend to carry out this virtual plan. Nor does the fox expect the rooster to carry it out. Nevertheless, both characters talk about it as if it were real. The fox, for instance, says, "Will you come to my house for breakfast?" thus pointing to the virtual plan. The rooster accepts the virtual plan, or at least appears to do so, when he says, "Yes, thank you, I will come," and then he suggests a modification when he says, "if my friend may come too." The fox accepts the modification, saying, "Oh yes, I will ask your friend. Where is he?" Thus we have a situation in which the characters discuss and modify a plan which neither expects to be carried out.

Virtual plans are common in stories. Hansel and Gretel's parents use the virtual plan of ordinary wood fetching to pursue their real plan of getting rid of the kids. Meanwhile, Hansel and Gretel use a similar virtual plan to achieve their real goal of returning home. In fact, the following outline appears to be a good model for a large class of children's stories. It applies, for instance, to "Hansel and Gretel" and "The Dog and the Cock." It defines a kind of deception wherein characters act on the basis of real plans, but pretend to act on the basis of virtual plans.

Outline of a Typical Story Containing

a Virtual Plan

1. A has a problem which suggests a goal that is in conflict with

    a goal of B.

2. A realizes that B's normal actions (or inactions), i.e., B's real

    plan, will not help him to achieve his goal.

3. A further realizes that B will not alter B's plan to suit A's goals.

4. A therefore puts forth a virtual plan either to conceal A's real plan

    or to entice B into doing something B would not otherwise have done.

5. B responds to the virtual plan. In some cases B falls for the trap,

    e.g., in Aesop's fable of "The Fox and the Crow" the crow sings in

    response to flattery and drops a piece of meat. In other cases, B

    sees through the virtual plan to A's real plan, then pretends to go

    along with the virtual plan, or puts forth B's own virtual plan.

6. Actions proceed, but each action has alternate simultaneous interpre-

    tations, as part of the virtual plans and as part of the real plans.

7. At some point the virtual plan is discovered, or uncovered, and the

    story (or episode) draws to a close.

e. Critical Beliefs

One of the most interesting results of our analyses in terms of inter-
acting plans has been the realization of the tightly interwoven character
of the plans representations. Any representation of "meaning" can be
viewed as an arbitrary and unsatisfying abstractions from the "whole," but
plans seem even less divisible than other facets of meaning. It is much
more difficult, for example, to change one small part of a plan's represen-
tation without producing rippling effects throughout the representation.

A consequence of this holistic property of plans is that a single be-
lief can assume tremendous importance. In "The Dog and the Cock", the
reader's belief that the Cock believes that foxes like to eat cocks appears
to be a critical belief for the building of the typical adult interpreta-
tion of the story. Some children do not seem to have this belief and build
a different interpretation in which the Cock is an unwitting potential vic-
tim of the Fox, who is saved through no effort of his own by the Dog.

The latter interpretation is internally consistent, and it matches
the story as well as the typical adult interpretation. Is it therefore
also correct? How many different interpretations are there? We may not
be able to answer those questions, but we can observe that the one criti-
cal belief has had significant ramifications for the interpretation. Con-
sider how readers with the two interpretations would answer the following
questions:

1. Did the Cock trick the Fox?

2. Did the Fox trick the Cock?

3. Was the Cock smart?

4. Did the Cock think that the Fox liked his singing?

5. Was the Cock happy with what the Dog did?

The notion of critical beliefs seems worth pursuing. It may account for some differences in interpretations due to cultural variation found among readers. It also needs to be considered when we think of testing for comprehension. Finally, it shows one more way in which the things the reader brings to the text are as crucial to understanding as what is "in" in the text.

f. Complexities -- Easy vs. Hard Texts

A formal analysis demonstrates that even apparently simple stories may require complex plans representations. A reader needs to be able to induce plans from the often sketchy statements of actions and intentions. He or she must then be able to use the induced plans to connect events. There are a number of specific abilities a reader would need in order to understand plans in this way. We do not know that these abilities are a major cause of reading comprehension difficulties, or even that they form a complete or well-defined set of skills with respect to understanding plans. Rather, they point to areas that might be worth investigating.

Among the complexities are the following (see Bruce & Newman, 1978, for a more detailed discussion):

1) Changes in Plans. Plans in a story can remain fairly constant, like Babar's plan to retrieve his crown, or may change in response to events. The number and magnitude of changes may be a source of difficulty.

2) <u>Size of plan</u>. Plans vary in their inherent complexity. A plan may involve a long sequence of acts or may be accomplished by a single act.

3) <u>Embeddings of beliefs</u>. Whenever a belief is about another person's beliefs, one must be able to shift point of view. Sometimes a story, e.g., "Hansel and Gretel," requires multiple shifts of point of view, to beliefs about beliefs, etc.

4) <u>Embeddings of intention</u>. Similarly, intentions may be embedded. For example, Hansel and Gretel's parents intend the children to have the intention of following them into the woods.

5) <u>Embeddings of plans</u>. A consequence of the embeddings of beliefs and intentions is that one's plan can be defined with reference to other plans, and those plans to yet other plans. Hansel's stepmother, for instance, tries to block Hansel's attempt to block her plan to abandon Hansel and Gretel.

6) <u>Degrees of interaction</u>. When there are multiple actors in a story, their plans can be more or less interconnected. Hansel and Gretel's plans are intertwined with their parents' plans. Each is trying to respond to the others and to get the others to do an act in a particular way. In other stories there may be only one character, or characters' plans may not interact as tightly.

7) <u>Deception</u>. A story that involves deception, e.g., "Hansel and Gretel," is inherently more complex than one that does not.

8) <u>Conflicts</u>. The number and types of conflicts among plans in a story may also be a source of complexity. In a general sense, we can view actions as attempts to reduce conflicts among plans. For example, there is

a potential conflict between a plan I believe you to have and the plan I want you to have. There is also a potential conflict between a plan I want you to believe I have, and the plan I believe you believe I have. It is not necessarily the case, however, that plans of any type can conflict with plans of any other type. In fact, the identification of types of plans leads us to an identification of types of conflicts that can arise among plans in interactive situations. This suggests a number of questions about people's recognition of and response to such conflict situations. For example:

a. To what extent do people of different ages recognize embedded plans?

b. How deep do these embeddings go?

c. Do the potential conflicts actually arise in all situations?

d. How sensitive are people to the ability of those they are interacting with to perceive such embeddings and conflicts?

e. How are the virtual plans and the conflicts among them signaled in text?

f. What are the consequences of one's understanding of different levels of interpretation of virtual plans?

9) <u>Maintaining different points of view</u>. Having to maintain different points of view, e.g., that one character believes X where another believes not-X, imposes demands on the reader. In addition to the levels of embedding mentioned above, there may be problems associated with maintaining a large number of differing beliefs or maintaining any differences for an extended period.

10) Cultural presuppositions. Stories that involve beliefs about
character types or simply facts about the physical world may place differ-
ential demands on readers depending on their experiences prior to reading.

11) Beliefs outside of "shared belief space." In a normal episode
most beliefs are "shared" among the participants, meaning not only that
they both believe, but that they believe that the other believes, and that
the other believes that they believe. The reader can then assume that all
knowledge is transparent to all. Often, though, one must assume that there
are beliefs outside of the shared space, not necessarily conflicting
beliefs, but beliefs that are not known to one or more characters.

12) Inference. The number of extra beliefs needed and the amount of
deduction required to link together actions in a story is also an indica-
tion of possible difficulties for the reader./

13) Explicitness of plans. Texts vary in the degree to which they are
explicit about the plans and intentions of characters in the story. Stories
are more difficult when the reader has to infer plan structures from the
simple statements of actions.

14) Act hierarchies. An important aspect of interacting plans is that
people develop them and carry them out in the context of their perceptions
of others' actions. The same action can be viewed at various levels or
clumped together with other actions. There can be many levels of concep-
tualization for the same act or sequences of acts.

g. Conclusion--Implications for Education

It would be presumptuous to suggest revisions in education on the basis
of an interactive plans analysis of a handful of children's texts. After

all, the variation among texts could be greater than that suggested here;
the complexities that have been identified have only tentatively been shown
to cause specific problems in comprehension; and the method of analysis is
itself new and untested. Nevertheless, there are clear, albeit tentative,
signs that have emerged from the analyses we have done.

The comments below are separated into two overlapping categories:
issues related to texts and issues related to learning to read. The first
set applies to decisions about selection of texts for beginning readers,
for tests of comprehension, and for workbooks and similar "school texts,"
The second set discusses teaching methods, testing, and the development of
reading skills.

1) <u>Text selection</u>. First, with regard to text characteristics, the
interactive plans analyses that we have done illuminate a world of phenomena
that are implicitly ignored in the design and selection of texts for use in
school workbooks, tests, primers, and textbooks. Stories such as "Indian
Occupation" are commonly used, although they are very poor as literary
texts, because they offer an opportunity to teach such things as the "oa"
sound. But what effects might frequent exposure to such texts have?

For the fortunate child who has books at home and parents who support
and encourage literary explorations, there may be no harm done and the drill
on a specific skill may be useful. But such a child is not the one with the
greatest reading problems.

Think for a moment of the child who has limited reading experiences out-
side of school, who has few picture books, and who does not hear stories being
read. In the early grades, he or she encounters a series of texts that stress

decoding skills. The texts often sacrifice the story line under the assump-
tion that component skills need to be taught independently. Thus, it is
assumed, story structure can be taught when its time comes; there is no need
to demand high quality stories when one is teaching decoding. Later in
school there are expository texts to read. It is assumed that the child is
already a reader, even though critical reading has been rarely taught, or
even encouraged. The skills the child is supposed to have learned are just
to be "applied" in learning new subject matter. Throughout the elementary
grades, the literary diet of the child fails to exercise the skills of
critical reading. Often the result is that motivation suffers as well as
skill development.

If an interactive plans analysis shows nothing else, it still demon-
strates that understanding even a "simple" fairy tale requires sophisticated
skills. Where are these skills to be learned, if not through reading (or
being read) good texts?

It is only partly facetious to propose a text quality hierarchy of the
following kinds:

  a. Texts never seen in school.

  b. Texts allowed when the regular work is done.

  c. Texts read for a purpose other than learning to read.

  d. Texts used to teach reading.

  e. Texts used to test reading ability.

  f. Texts used to teach specific component skills (often used in
     remedial classes).

A reader who gets enough of types a and b will learn to enjoy reading, will
learn that there is a point to reading, and will learn the skills necessary to

read with understanding. These three types of learning exist in positive

feedback loops, as shown in Figure 10.
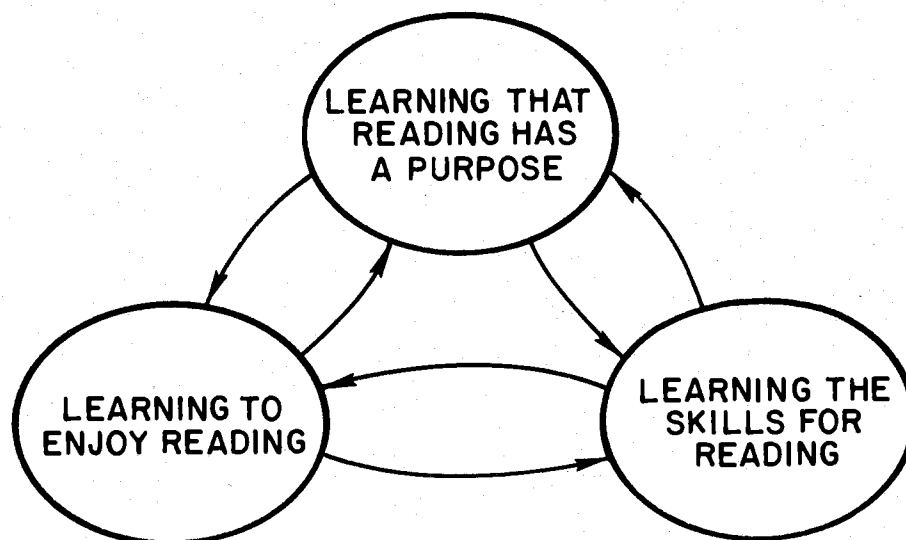
Figure 10
Learning to read

But the child who only sees texts at levels c and d, or worse never enters the

loops. He or she is then tested at level e and punished with level f texts

for failure to perform on the test.

The text quality issue interacts with another, complexity. While it is

important to realize that reading is a complex skill, we must not assume that

"simpler is better" with regards to text selection. Complexity is multidimen-

sional; the best text may be one that challenges the reader on a few dimensions

and allows easy success on others. The attempt to produce a single number mea-

sure of complexity or difficulty is probably misguided.

Furthermore, when one moves beyond the orthographic and lexical levels of

analysis one finds more and more a tendency for texts to be understandable in

different ways. We suspect that few three-year-olds, upon hearing "Hansel and

Gretel," would understand it as an adult would. Yet their understanding, though possibly limited, is not wrong in the sense that saying "cat" to the letters "D - O - G" would be. It is a characteristic of interacting plans and, we think, in general of the phenomena of literature that one can uncover multiple meanings, no one of which is wrong, or even unsatisfying. The best approach may be to give children the best literature (expository or narrative) that we can find, letting the child, rather than some arbitrary formula, decide whether it is too complex.

2) <u>Teaching reading</u>. What do interactive plans analyses tell us about learning to read? One point was broached in the previous section. Understanding plans in stories is a complex task that may require years of exposure to high quality texts to learn. Consequently, we should expect children and adults to understand stories in different ways, simply because they have had varying amounts of experience. It would not be surprising to find examples of understanding at each of the following levels to be an indicator of experience with reading:

a. Isolated sentence understanding -- each sentence is understood but connections are not made.

b. Islands of understanding -- local connections among sentences are made but no overall pattern is seen.

c. Limited plans understanding -- basic plans are comprehended, but not interacting plans.

d. Embedded plans understanding -- full understanding of the interactions among plans of characters in a story.

An interacting plans analysis also gives some guides for our expectations about developing readers. First, the complexity of plans means that

readers may understand in different ways, yet still be reading, and hence, learning to be better readers. Second, the importance of "critical beliefs" means that readers with different backgrounds may build divergent interpretations of the same texts. Both of these points need to be considered seriously when we think of what it means to test comprehension skills.

3. Story Grammar Analysis

This section has two goals. The first is to illustrate how a specific set of predictions about story memory can be made, using a story grammar analysis. The second is to raise some critical issues about story comprehension which have not been directly addressed.

In the past few years, it has become increasingly apparent that models of single word or sentence comprehension cannot account for many of the important factors affecting the comprehension of discourse material. Although theories of discourse comprehension must eventually explain how these smaller units influence the comprehension of an entire passage, an approach describing the semantic relationships between sentences is necessary.

In an attempt to extend Bartlett's (1932) work on story memory, Propp's (1958) morphology of the folktale, and Rumelhart's (1975) initial scheme for stories, several story grammars have been constructed to describe the structural basis of story understanding. A major theoretical assumption of these grammars specifies that memory for stories is a constructive process, resulting from the interaction between incoming information and pre-existing cognitive structures, containing knowledge about the generic characteristics

of stories. These structures or schemata, defined as a set of rewrite rules,

influence the way in which a listener will break down incoming story infor-

mation into its component parts. Thus, schemata aid the listener by speci-

fying the types of information and the types of logical relations which should

occur at various points in a story. It is then assumed that the listener can

determine whether any necessary information has been omitted or whether the

correct logical relations have been used to connect the various story components.

Thus, the major assumption underlying the description of a story schema is

that comprehension of a story involves the use of an ideal story structure

to reorganize, represent, and retrieve incoming information. When text struc-

tures do not conform to the rules specified by a story schema, then readers

or listeners will attempt to transform the incoming information so that a

representation adhering more to the structure of an ideal story schema can

be constructed.

In several studies on story comprehension, Stein and Glenn have described

a story schema in detail and have presented evidence to support several hypo-

theses concerning the validity of a story grammar (Stein, 1978; Stein & Glenn,

1977; Stein & Nezworski, 1978). These results will be summarized by pre-

senting a description of the types of story recall that should occur if a

narrative such as Babar Loses His Crown were to be presented to adults and

children. In order to understand how a schema interacts with incoming story

information, a brief description of the process of analyzing a simple story

into its component parts is presented. Although BABAR is a complex multiple

episode story, the example of a simple story breakdown is representative

of the process involved in organizing information from complex stories such as BABAR.

According to the Stein and Glenn (1977) grammar, the primary unit of analysis in a story is the category, and several categories occur within a story structure. Each category refers to a specific type of information, and serves a different function in the story. Normally, each sentence in a story can be classified into a particular category. However, the sentence is not the critical variable defining category membership. There are instances in which the initial part of the sentence belongs to one category, and the latter part to a second category, depending upon the functional role each portion plays in the story. It should also be emphasized that a story category can contain one statement or several such statements.

A story structure can be described in terms of a tree diagram which is a hierarchical network of story categories and the logical relations which connect them. The initial division of a story consists of two parts: a setting plus an episode structure. The setting begins the story with the introduction of a protagonist and normally includes information about the social, physical, or temporal context pertaining to the development of the episode. The setting is not part of the episode, as it is not directly related to the subsequent behavioral sequence described in the episode. However, information in the setting category may constrain the possible types of behavioral sequences which then occur.

The remaining story information in the episode consists of a sequence of five categories: initiating event, internal response, attempt, consequence, and reaction. The initiating event category contains some type of

event or action which marks a change in the story environment. The major

function of this change is to evoke some type of response from the pro-

tagonist which is defined as the internal response category. Internal re-

sponses can include goals, affective states, and cognitions, and serve to

motivate a character's subsequent overt behavior. Actions which describe

this overt behavior are defined as attempts. A character's attempts then

result in the direct consequence of a character's actions, marking the

attainment or non-attainment of the character's goal. The final category

is the reaction which can include a character's internal response to the

consequence or broader consequences caused by the goal attainment. If the

relationship between the setting and episode is ignored, it is apparent

that each category logically follows the preceding one. Furthermore,

according to the grammar, these categories always occur in a specific

temporal order.

There are several factors which alert a reader or listener to the

fact that one category has ended and another one begun. Temporal markers

such as "One day," "Suddenly," "Finally," etc. often signal the beginning

of a new category, facilitating the breakdown of stories into components.

The semantic content of a statement and the relationships among statements,

however, are just as important in determining the division among categories.

From the previous description, it is evident that certain types of

information are always contained in the internal representation of an epi-

sode and that the temporal order of category information and the logical

connections between categories are also critical components of a story

schema. By using the rules described in the grammar, a set of predictions has been made about the nature of story memory, with specific reference to story recall. Although recall does not guarantee an isomorphic correspondence to the underlying representation of story information, it does allow an initial assessment of the types of story information thought to be most critical in maintaining the semantic cohesiveness of the text of a story. Therefore, the next two sections will discuss predictions concerning the types of information most frequently recalled, the types of new information which will be added to a story, and the order in which story information will be recalled.

a. Probability of Statements Occurring in Recall

In past studies, it has been argued that several factors influence the probability of recalling individual story statements. A story statement, in most instances, is equivalent to a single sentence in the text of a story. The sentence, however, is not necessarily the critical component for defining whether information can be classified into one or two statements. Rather is is the function of the information within the context of a story that is the critical determinant of the unit of analysis. Stories are basically concerned with goal-oriented behavior and ideally consist of a sequence of statements directly related to the attainment of the goal. Therefore, the types of logical relationships existing among story statements in an episode are the critical factors in predicting the saliency of individual statements in recall. In the story studies, accurate recall is defined as the production of statements containing an extremely close correspondence

to the semantic content of the original story material. The syntactic form of story memory is <u>less</u> important. If the relationships among statements are directly causal in nature and are related to the character's major goal attainment, then they have a high probability of being recalled. In recall, then, certain story statements assume a more important role than other statements.

Two additional factors are important in predicting the saliency of each story statement. The first factor concerns the semantic content of the statement. Although two statements may be causally related to one another, the information in the first statement may directly imply the type of information in the second statement. In this situation, the recall of the second statement becomes unnecessary or redundant. A second factor concerns how well a particular story statement matches the type of knowledge acquired about the specific sequence of events being presented. Often two statements in a story will again be causally related to one another, but the listener will recall a statement that is an <u>integration</u> of both statements, or a statement which contains information from which the actual story statements could be inferred. Thus, the semantic content of a statement, as well as the type of relations among statements, affect the probability of recalling individual statements.

In past studies it has been found that certain categories of information were better recalled than other categories. The categories most frequently recalled were major setting statements (characterized by the introduction of the protagonist), initiating events, and direct consequence statements. Attempts were in the middle of the frequency distribution, while

Internal responses, reactions, and descriptive settings were infrequently recalled. These results seem to suggest that statements within these latter categories are either semantically redundant or not directly related to the protagonist's attempt to attain a specific goal.

The saliency of story information is also related to the organization of story information into higher order units. The episode is the main psychological unit in a story structure. Just as there are different types of relationships among statements within an episode, there are also different types of relations linking the episodes of a story. The relationships among episodes also play a critical role in determining whether story statements will be recalled. In many stories, such as BABAR, there is one overriding goal stated in the first episode of the story. The remainder of the story, then, consists of a number of episodes containing subgoals that are directly related to the protagonist's desire to attain the major goal. If an episode contains subgoals directly related to the major goal, it should be well recalled. However, there may be episodes contained in a story which have only an indirect relationship or no relationship to the major goal. Because these episodes are "empty" in the sense of being unrelated to the goal, they serve little purpose and are readily forgotten.

b. Temporal Organization of Stories

The next set of predictions derivable from story grammar analysis concerns the variables which regulate the ability of a listener or reader to recall the correct temporal sequence of a given story text. Two categories of results will be given.

1) <u>Within-episode sequences.</u> If the temporal organization of a story

text corresponds to the structure described by the grammar, subjects will

have little difficulty organizing incoming information and recalling the

temporal sequence of a story text. The data from several studies strongly

support this result. Even children as young as four can recall the current

temporal sequence of a story text. However, when the text sequence diverges

from or violates the order specified by an ideal form, listeners have dif-

ficulty maintaining the exact order of the text and reorganize the story

text in several different ways. The reorganization of the text is repre-

sentative of a tendency to recall information in an idealized form corre-

sponding more to the structure described in the grammar than to the sequence

given in the text material.

The type of reorganization that subjects will use during recall depends

upon: 1) what type of story information occurs out of sequence in the text

structure; 2) the text distance of the information from the hypothetical

location prescribed for it in an ideal structure; and 3) the relationship of

the moved information to the new surrounding story information. If the story

sequence in a text is altered by simply reversing the positions of two ad-

jacent statements, subjects recall the story in one of two ways: they either

reverse the two statements so that the order resembles that described in an

ideal structure or they insert a causal or temporal connector to signal the

fact that an inversion has occurred. When the information in an experimental

text is moved more than one location away from its "normal" position in an

ideal sequence, different strategies are used. In this case, subjects either

interpolate extra material to connect the moved information more

appropriately to surrounding information or they delete the moved information altogether.

2) Episode organization. In many stories, there are two or three episodes which are sequenced in an arbitrary fashion such that there is no a priori reason that one episode should occur before another episode. There is no direct causal relationship linking the two episodes. In this case, the probability of a subject's maintaining the correct temporal order of the episodes in recall decreases markedly. Thus, the type of connection linking episodes not only predicts whether an episode will be recalled, but also predicts the order in which episodes are recalled.

c. Current Problems in Studying Story Comprehension

Although preliminary results are very promising for the use of a story grammar to understand parts of the comprehension process, the important work in this field is in an initial stage. There are several issues which still need much more investigation.

1) Encoding. To date, most of the studies completed on stories have used only recall procedures. While recall is important in assessing memory in terms of retrieval processes, this type of methodology does not directly answer questions about the encoding process or the actual process of representation. Again, one of the major assumptions underlying memory for stories is the proposed interaction between incoming information and pre-existing operations and knowledge structures; it is not clear at the present time how the interaction of these variables differs during encoding and retrieval. For example, Stein and Nezworski (1978) have shown that subjects recall stories in a very specific type of temporal sequence and will

transform incoming information so that the order of output matches the order described in an ideal story structure. However, these transformations may occur only as a function of the constraints placed on working memory during retrieval. The underlying representation of the story may be more complex and more representative of the pattern of incoming information. Both Mandler and Johnson (1977) and Stein and Nezworski (1978) have stated that although there may be similarities in the process of using schemata during encoding and retrieval, there are also significant differences.

One method of showing the differences between encoding and retrieval processes is the use of recognition and probe procedures. Stein and Glenn (1977) have already demonstrated that certain types of story information (internal responses), infrequently recalled by children, are very accurately encoded (when probe techniques are used to assess comprehension). Stein and Nezworski (1978) have also shown that although subjects re-order stories containing violations of an ideal form, they are aware of the types of violation occurring in the structure of a text. Thus, during the process of encoding stories subjects incorporate much more and different information about the text structure than they are able to retrieve.

2) Inferences. The difference between the encoding and retrieval of stories also raises questions about the actual structure of the representation of stories. One of the more important issues related to the comprehension of stories concerns the types of inferences made during the process of encoding and representation. Story information often deals with moral dilemmas where children continually make inferences about the acceptability

of a character's goals, plans, and attempts to attain the goal. That is, inferences are spontaneously made about whether or not a set of actions are good or should have been performed and whether or not the character is seen as good or bad. Inferences are also made about a character's personality traits, affective states, and perceptions about some of the story events.

Although stories sometimes explicitly state this type of information, more often than not the reader must make inferences about this information. Because these inferences are related to the way in which children apply story information to their own problems and behavior, it becomes critical to begin a more detailed investigation of this type of information.

3) Semantics. The more general problem with existing story grammars is that they do not provide a method for understanding the resulting representation of the specific semantic concept of a story. Although general predictions in the Stein and Glenn grammar were made about information salience in recall, based upon the relational structure created among story statements, these predictions do not concern the specific content of the story material. The development of a more encompassing model is necessary, laying out specific predictions about the recallability of each specific story statement. Although we can show that causal relations among statements and their relation to the character's goal are important, these factors do not allow enough specificity about recall.

A final issue which needs investigation concerns the changes which occur in story structures as a function of age. In one study on story production (Stein & Glenn, 1977), it has been shown that story length and

the complexity of the story structure change as a function of age. Young children do not produce certain types of stories (e.g., interactive character episodes, complex goal structures, etc.). It is now necessary to determine how these different structural characteristics affect comprehension.

4. Analysis of Expository Structure

The types of higher order cognitive structures used during expository comprehension have not been described in as much detail as those for stories. Although Meyer (1975) has specified some of the types of information that occur in expositions, she has not proposed a general set of rewrite rules which specify the types of information that must occur in certain texts or the relationships which must occur among the various types of information. Because there has not been much effort geared to describing the internal representation of these texts, it is difficult to make an explicit set of general predictions concerning memory for and comprehension of exposition.

a. Expository Description

Although the structure of an exposition has many unique components when compared to story structures, there are many similarities between the two types of structures. In order to illustrate these similarities, we will take exposition and compare it to the structure of a story. First, in a "Definition Exposition," there is a major purpose inherent in the text. That is, the purpose of texts which define a notion is to transmit specific information about it to a reader. There are several additional assumptions which need to be made if we accept the initial one.

When information is transmitted to a listener, we must also assume that there is a purpose for the transmission of such information. Normally,

a reader has a specific set of reasons for wanting to learn about something.
He is not only concerned about the unique descriptive characteristics of
an object, but he is also concerned about the functions and uses of an ob-
ject. These factors cannot be ignored in describing the characteristics of
an exposition. There is purpose, directionality, and a cohesive structure
underlying a definition exposition.

In the first part of a definition exposition, labelled the generic de-
scription category, two types of information could be included:

1) the defining features of the object

2) a similarity continuum, defining the object by comparing

   its features to objects presumed to be already familiar

   to the reader.

The defining features of the object are a critical part of the text and can-
not be omitted. The representation is defined as a list of features which
characterizes the object and differentiates it from all other objects.

When the purpose of exposition is to provide new information, the second
category of information, the similarity continuum, is often introduced. One
of the major purposes of this category is to enable the reader to use pre-
existing knowledge structures to encode and construct a representation of the
new information. By providing information which shows both the similarities
and differences to knowledge already acquired about other objects, a new re-
presentation can be constructed which incorporates some information from
other structures but not all of the information. Thus, not surprisingly,
this portion of our hypothesis bears similarities to the processes and re-
presentation proposed for the comprehension of a metaphor (Miller, 1979).

The types of structures proposed for encoding expository information are also similar to those proposed in Rumelhart and Ortony (1976).

The inclusion of the second category of information may be dependent upon how much information the reader already has acquired about similar types of objects. If the subject's pre-existing knowledge is extensive, then inclusion of this category in the text may be necessary so that the definition features attributed to the object can be instantiated correctly. If similarity-differentiation information is not included, subjects may have a tendency to overgeneralize from the knowledge they already have, and incorrect features may be attributed to the object. If pre-existing knowledge is not extensive, then the defining features may be remembered without this type of information.

The relationship or logical link between the first two categories can be characterized as an enabling relationship. The description of the defining features ENABLES the reader to initiate a search for knowledge or information similar to or different from the specific entity being described.

The third category, functional consequence, is initiated as a RESULT of encoding specific features about an object. In the DESERT text, the types of consequences which follow from the object description, are negative in nature, signifying the harm that can occur to all forms of life without the existence of water. However, in many texts, the type of object being defined can also be conceptualized as first having positive consequences associated with its function. The type of consequences first introduced in an exposition should be a function of the knowledge already acquired about

similar types of objects and their defining attributes. These predictions assume that readers automatically make inferences about either positive or negative functions of an object. That is, there is a valence implied in the description of generic attributes, and the quality of this valence will determine whether positive or negative consequences are stated first.

If negative consequence category is given, the category following this should be a PREVENTATIVE ACTION category. The inclusion of this category assumes that the listener has specific sets of expectations such that he desires to know about a specific type of end state when acquiring new information about an object. This assumption is similar to that proposed by Heider and discussed by Abelson, Aronson, McGuire, Newcomb, Rosenberg, and Tannenbaum (1968). Thus, the inclusion of negative consequences should cause the reader to _infer_ preventative actions to neutralize a negative consequence state.

The fifth category, positive or neutral consequences of an object function, is the last category in this structure. The relationship of this category to the previous one is an AND relation. It is assumed that the statement of negative consequence and preventative actions does not CAUSE or ENABLE the statement of positive consequences. Thus the relation between the two consequence categories must be defined as an AND relation. It should be noted, however, that the structure of knowledge about objects probably contains information about _both_ types of consequences; and there may be an underlying set of relations, not specified here, which allows the reader to make spontaneous connections between the two opposite types of states.

B. Discourse Properties

This section describes a number of text-level discourse properties which have characteristics which may affect reading comprehension.

Part 1 consists of a general discussion of information structures which are posited as scaffolding for the erection of a model of the world of the discourse in a representation of the comprehension process. Topic-comment organization and story grammar organization are both treated, and discussion is included of their relation to each other and of their relevance to different kinds of texts.

Part 2 is a brief survey of some of the rhetorical devices used in texts to emphasize and de-emphasize, or foreground and background, various kinds of information. A variety of syntactic devices are discussed which reflect the narrator's or a character's point of view, or which reflect the narrator's attempt to influence the reader's perception of events described.

Part 3 contains discussion of three major text-connective devices: scene-setting devices, conjunctions, and anaphoric devices.

1. Information Structures

There has been a great deal of interest recently in the notion of information conveyed by a sentence, how information is conveyed, what its form is, and whether the structure of a sentence or discourse is dependent on or reflects the information communicated. It has been postulated that comprehension of text depends crucially on the ability of the reader to match up the text with an appropriate organization within which to interpret it. In this section we examine a number of the terms used in the description of information structure schemata, and evaluate the feasibility and utility of applying them to a complete text such as BABAR or DESERT.

Let us begin with the notion <u>topic</u>. Some authors use this term to mean <u>discourse topic</u>, or what a discourse is about. Other authors use it to mean <u>sentence topic</u>, or what a sentence is about. Topics are sometimes considered to be unexpressed <u>propositions</u> which characterize "aboutness" of a discourse (see Keenan & Schieffelin, 1976). That is, throughout a discourse, topics are established, and any given sentence will be related to the current discourse topic or establish a new topic. There do not seem to be any good tests for this kind of discourse topic, and, as we will see below in a partial analysis of topic in BABAR and DESERT, the notion "discourse topic" still needs quite a bit of refining. According to other scholars, topics are syntactic elements. These topics are either surface constituents in some languages (see Li & Thompson, 1976), so that the basic surface structure of sentences is [topic-comment], as in Lahu, Lisu, and Chinese, or they appear as nodes in a semantic representation that may or may not be syntactically realized (see Gundel, 1977).

Another term used by some text-structure theoreticians is <u>theme</u>. This is used variously to mean 1) the point of departure of a sentence, i.e., the element that comes first in the sentence (see Halliday, 1967); or 2) those elements in the sentence that are already within the hearer's or reader's sphere of knowledge, i.e., information that has already been introduced to the reader or hearer by the writer or speaker (see Daneš, 1974).

Another set of terms is <u>given</u> and <u>new information</u>. <u>Given information</u> is used to mean the information the speaker believes to be present in the (short-term) consciousness of the hearer (see Chafe, 1976), or what the speaker believes to be recoverable from previous discourse (see Halliday, 1967).

Chafe (1976) proposes as a distinct term, <u>known information</u>, which he takes to be those elements that are subject to definitization in a sentence. Kantor (1977) suggests a concept of <u>degrees of activatedness</u> of information in the consciousness of the hearer or reader to explain the degrees of comprehensibility of various referential expressions, e.g., pronominal, demonstrative, definite, definite with added descriptors.

A major problem with almost all the studies referred to above is that they give no methodology for the the analysis of texts in general. Indeed, the analyses used to exemplify the theories are usually restricted to critical texts composed of between two and six sentences. As a result, we find the various constructs to be difficult to apply to actual texts like BABAR and DESERT.

To illustrate, let us take the notion of discourse topic. The notion itself seems quite transparent, i.e., that the discourse is at any point about something, someone, or some event. But notice what happens when we try to apply this notion to our texts:

Title: <u>Babar Loses His Crown</u>

1. The Babar family is going to Paris.

   (What is the discourse topic? Perhaps that the family is going to Paris.)

2. Everyone is packing bags.

   (This sentence describes what is going on in the illustration. Perhaps the topic is what the family does to prepare to go to Paris. But this does not seem very satisfying.)

3. Here are the children--Pom, Flora, and Alexander.

4. Here are Cousin Arthur and his friend Zephir, the monkey.

   (Introduction of actors; the discourse topic might be charac-

   terized as 'the Babar family' or 'who everyone is'.)

5. Queen Celeste and King Babar pack their crowns.

   (More introduction; the topic is 'who everyone is and what they

   are doing towards getting ready for the trip'.)

6. Babar puts his crown in a little red bag.

   (Topic is 'what Babar is doing towards preparing for the trip'.

   The most interesting fact about this sequence is that it focusses

   on Babar's crown and the little red bag, which becomes very im-

   portant to the story later on. Thus, the function of the sen-

   tence, i.e., to foreshadow the future action is what is important

   here. The topic is not terribly revealing.)

7. Now the Babar family is on the train.

   (Topic is perhaps 'the Babar family's trip'.)

8. The train is coming into Paris.

   (Topic is still 'the trip to Paris', with perhaps a subtopic

   'travelling by train'.)

9. "I will show you everything," says Babar.

10. "You will love Paris."

    (What is the topic here? Perhaps still 'the train trip to Paris'?)

11. Now they are off the train with all their bags.

    (Sets a new scene; perhaps topic is still 'the trip to Paris'.)

12. The Babar family is waiting for a taxi.

    (This is a continuation of the action; the topic might be charac-

    terized as 'the family in Paris'.)

We could go on through the entire story with these rough descriptions of what the sentences are about. However, all we would have is what the sentences are about, according to our intuitions, without any methodology to tell us how to direct such intuitions. Furthermore, it would not tell us anything revealing about how this notion contributes to the process of constructing a coherent interpretation of a text.

Rather than a strictly intuitive approach, we could perhaps employ a methodology of "filling in" the missing information which, it might be maintained, will provide the topical connection between sentences. For example, we might try to turn the story into one or more long sentences:

1.   The Babar family is going to Paris (and toward this end,)

2.   everyone is packing bags.

This will become awkward to do, however, when we encounter direct discourse, as shown here:

7.   Now the Babar family is on the train, (and as for this train,)

8.   the train is coming into Paris, (and with respect to Paris,)

9-10. "I will show you everything," says Babar. "You will love Paris."

Now while such marking of the relationships is possible, we are faced with at least two problems. First, we must have a set of criteria whereby the marking of the relationship is done by some specified procedure, so that any investigator will arrive at the same relational phrase. Secondly, we will want to be sure that the relationships that we are marking are relationships of topic, rather than something else, say, logical relationships between two sentences. Finally, we must also be able to indicate that sentences like Babar's direct discourse in 9 and 10 are somehow subsidiary to the main topic

of 'the family going to and arriving in Paris'. For this we need at least a hierarchy of topics and subtopics. We lack a methodology for performing such an analysis at present.

An analysis of discourse topic in an expository source like DESERT is somewhat more intuitively satisfying, as shown in the following fragment:

1. What is a desert?

   (This sentence sets the major topic, 'the desert'.)

2. You think at once of a place without water.

3. You would think that the desert does not have enough rain for

   things to grow.

   (The topic here is 'what is generally felt about deserts'. The

   author is using these sentences as a device for introducing his

   main thesis, which is . . .)

4. Yet there is life in every desert.

   (This sentence quite clearly is the topic sentence of the text.)

5. Plants, animals, and people live there.

   (This sentence has a topic 'there is life in every desert'. It

   sets the structure of the text to come, i.e., it introduces the

   subtopics of 'plants, animals, and people in the desert'. This

   sentence is a statement of support for sentence 4.)

6. How can this be?

7. It is a wonderful story.

   (These sentences clearly relate to the topic 'there is life in

   the desert' but how, exactly, we have not yet been able to charac-

   terize.)

NEW PARAGRAPH

8. In the desert it may not rain for five years.

9. Then one day a storm comes.

10. A heavy rain falls.

    (All three of these sentences should probably be taken as one

    topical unit, setting the scene for a new subtopic, 'plants in

    the desert.')

11. All at once the desert is covered with green grass.

    (Topic is 'how plants manage to live in the desert.')

12. You can see many small flowers.

    (Parallel to sentence 11.)

13. They grow very fast.

14. In one week they go from seed to flower and back to seed.

    (Elaboration on sentence 12; the subtopic here is 'flowers

    in the desert.')

15. Then the seed lies in the sand, perhaps to wait another five years

    for rain.

    (This sentence continues the action from 14; the topic is still

    'flowers in the desert,' but it also makes a back-reference to

    the beginning of the paragraph, sentence 8, to tie up the para-

    graph quite nicely.)

The analysis of DESERT in terms of topic is somewhat more satisfying,

for we can more readily identify what each of the sentences is about and

how they relate to the exposition as a whole. Does this mean that the

notion <u>topic</u> is applicable to descriptive discourse, but not to narrative

discourse? Possibly. Accepting an affirmative answer to the question means

that accepting topic as a linguistic construct entails that we take the

view that different genres or text-types are constructed of different kinds

of basic constructs. In either case, a non-intuitive methodology is still

necessary before this sort of analysis can be taken seriously.

It might be supposed that an analysis in terms of thematic (or given)

and rhematic (or new) information might be more insightful for narrative

texts such as BABAR. Some theories, especially those of the Prague School

linguists, e.g., Daneš (1974), propose that we can identify those elements

in a sentence that are old or given or thematic information as those ele-

ments that can be assumed to be known to the reader from previous context.

Following are the results of an attempt to apply this notion to the first

five sentence of BABAR. Old (known, given) information is represented in

braces.

Title: Babar Loses His Crown

1. The Babar family is going to Paris.

2. Everyone is packing bags.

(Here, it is debatable whether 'everyone' should be marked as

known information. If we make the assumption that the Babar

family consists of individuals, then we might say that one could

infer the existence of these individuals, and so, 'everyone' is

in some sense known to the reader. On the other hand, the reader

does not necessarily know that it will be the entire family that

is packing, and so the known information might be claimed to be

the semantically ellipsized 'of them.' Had the sentence been

'Every one of them is packing bags,' then we would have marked

'them' as known.)

3. Here are the children--Pom, Flora and Alexander.

4. Here are Cousin Arthur and {his} friend Zephir, the monkey.

    (What about the phrase 'the children'? Given 'family,' one can

    infer "possibility of children" at least. So, in a sense, "the

    children" might be characterized as 'known,' although there is

    no prior mention of them. In sentence 4, we can mark 'his' as

    known, since its referent has just been introduced.)

5. Queen Celeste and King Babar {pack} {their} crowns.

    ('Babar' is mentioned in the title of the story. It might appear

    that we should then mark this element as "old," but we can also

    make the case that having read the title, the reader does not

    necessarily know who Babar is, or what Babar is. Therefore, we

    do not mark this element as old information. The possessive 'their'

    is marked for the same reason as 'his' in sentence 4. The verb

    'pack' may be marked as some sort of old information, since the

    reader knows from sentence 2 that everyone is packing bags. But

    verbs, unlike nouns, do not refer. This immediately raises dif-

    ficulties of analysis, for the term "information" is a rather

    nebulous one. Does it refer to concepts, or propositions, or what?

    There is clearly some added cohesiveness to the story with the

    knowledge that Babar and Celeste are packing, since the reader

    knows that everyone is packing, so perhaps it is reasonable to

    consider the notion of packing as old information. But what is

    the nature of this old information? What does this claim commit

    us to saying about the representation of linguistic structure or

    about the psychological processes of the speaker?)

As we hope is evident, the characterization of terms like old or given

information and the methodology for their assignment to text elements is

entirely inadequate to the task of providing a reproduceable analysis of a

complete natural text. The whole notion of old information was proposed to

account for the fact that in a number of Slavic languages and other languages

of the world, word order that had been thought to be free (i.e., major sen-

tence constituents could be arranged in most any order) could be seen as con-

strained in discourse by the character of the information communicated accord

ing to the principle: 'old information first, new information last.' So,

for example, sentence 5 of our discourse when translated into a Slavic lang-

uage like Czech might have had the constituent order Verb-Subject-Object,

since 'pack' is termed "old information." But in a language like English,

where the restrictions on word order are much more stringent, it is not at

all clear what the concept of old or known information can contribute to our

understanding of the structure of sentences, the structure of texts, the

structure of English, or the structure of knowledge.

Perhaps if the notion of old information were clarified and refined, we

might find that the ratio of new information to old information in stories

should be lower for one level of reading ability than another, but with the

present state of the art, we find ourselves incapable of even beginning such

a study. Most linguists believe that there is something to the concepts of

topic and information structure, and so we would propose continuing research

in this area with an eye toward tightening up definitions and methodology.

2. <u>Rhetorical Devices</u>

a. <u>Introduction</u>

This section contains discussions of a number of syntactic devices exploited in texts to influence the way information contained in the text will be appreciated. Some of these, like the various (usually preposing) devices which in narrative texts indicate that the narrator ascribes particular importance to the events or individuals described, may help the reader to anticipate events or types of events, and reconstruct the structure of the text. Others, such as rhetorical questions and interjections like "Poor Babar!" which have forms which indicate that the narrator is addressing the reader personally, may have motivational functions. All of these speculated functions are amenable to experimental verification, the results of which could be applied almost immediately in the writing and editing of basal readers and other texts intended for reading instruction and reading improvement.

b. <u>Some Examples from a Narrative Text</u>

There are a number of places in BABAR where the author shifts from his standard detached and unbiased narration. Tagging each of these with labels would be premature, since the data to be discussed are mostly one of a kind. Further study of real and made-up texts needs to be made to see if any generalizations will obtain. Here, however, are some of the passages and some rough characterizations of them which identify devices meriting further attention.

> (93-98) But Babar can't eat. He is thinking about his
> crown. He needs it tonight. He must wear it to the
> opera. He fears his crown is gone forever.

The author is clearly speaking as if from inside the character's head here. The third and fourth sentences especially might well have been quotes, i.e., Babar thinks, "I need it tonight. I must wear it to the opera." Nowhere in BABAR is there any thought-quotation. This seems to come closest.

> (84-88) The Babar family circles around him. Babar says, "Ahem!" The man looks up. He is not the Mustache-man at all. "Oh, excuse us," says Babar . . .

The third sentence is interesting because of the phrase at all. Again, the narrator is speaking from the protagonists' point of view. The phrase at all is oral syntax, and reflects the expectations of the speaker. A more neutral description of the scene would be accomplished by something like He is not the Mustache-man or He turns out not to be the Mustache-man. Linguistic forms which reflect speaker's attitude are pervasive, even in English. Analysis of the rhetorical exploitation of such forms would be greatly aided by even a preliminary dictionary of such forms.

A final example:

(115-116) Poor Babar! His crown is lost again.

This is clearly a plea directly from the narrator for sympathy for Babar. Also, the use of the word 'lost' reflects his taking Babar's point of view in a subtle way. Previously, 'lost' was used only in direct quotation of Babar. Here, instead of describing Babar's thoughts, as he previously did in sentence 98, "He fears his crown is gone forever," he takes Babar's point of view, and simply speaks of the crown as missing, as if it were his own.

3. Connective Devices

This section contains discussions of three important elements which function to weave the threads of discourse in a text into a coherent entity.

These are 1) scene-setting sentences, which serve to mark the boundaries of units of text defined by physical location of characters; 2) conjunctions, which specify relations between the clauses they introduce and other clauses or larger chunks of text; and 3) anaphoric devices, which relate references to individuals within the text, subject to a set of often subtle constraints which interact in ways that are not yet entirely understood.

All of these devices may affect comprehension by making it more or less difficult for the reader to reconstruct the structure of the text, so that he or she will be able to perform such processing tasks as are necessary for a full appreciation of the text, e.g., forming expectations as to what will come next, picking out relevant details in what he or she is processing or has just processed, etc. Once it has been ascertained exactly how the devices reflect and signal information about structure and other discourse-level properties, it will be possible to evaluate texts with regard to them, and to establish how much of a factor they actually are in determining comprehension. This done, there may be implementable implications for the writing and editing of texts at all levels.

a. Scene Setting

In this section, we examine the syntactic and lexical markers of discourse structure which signal scene-setting sentences. The purpose of such sentences is to locate the new scene or episode in time or space. To see what kinds of generalizations could be made about the linguistic form of such sentences, we picked out those sentences in a narrative text (BABAR) which denoted a new location or time for the main characters, i.e., the Babar family. If the story described in detail how the family got to a particular

location or time, then any sentence mentioning this location or time was
not included. Following are the sentences that we identified as scene
setters on the basis of the contribution of their semantic content to the
development of the narrative:

7. Now the Babar family is on the train.

11. Now they are off the train with all their bags.

17. In their hotel room, Celeste opens all the bags.

32. So out they go, looking for the man with Babar's bag.

35. Now they are up in the Eiffel Tower.

40. A boat is ready to take off.

68. Arthur is standing in the sun to dry his clothes.

87. Now it is noon.

104. The Babar family is in two taxis.

137. Back in their hotel room, they say goodnight to the children.

143. They arrive at the big opera house.

It turns out that many of these sentences have one or more syntactic or lexi-
cal properties which make them natural candidates for a scene-setting func-
tion. Sentence 7, 11, 35, and 104, which locate the characters at a particu-
lar place, are all of the form Now Subject be Prep. Location. In addition,
sentence 32 contains a directional adverb or preposition which is brought to
the beginning of the sentence. Sentences 17 and 137 contain what might be
called thematic adverbials or locatives, preposed phrases in the sentences
which serve to locate the action in time. Sentences 68 and 143 contain the
verbs arrive and stand, which might well be classified as scene-setting verbs.
Notice that either of these verbs can occur after there, as in There stood

three men on the beach or There arrived a swarm of bees just as the summer
ended. Finally, we have sentence 87, Now it is noon, which explicitly sets
the time of the scene.

Each of these scene-setting sentences, then, appears to have one or more
markers characteristic of scene-setting. Further research is, of course,
required before we know whether sentences with these syntactic properties
necessarily serve a scene-setting function, whether what we have classified
as scene-setting has any psychological reality, and what other linguistic
forms may serve a scene-setting function.

b. Conjunctions

Conjunctions play a crucial role in providing the reader with informa-
tion both about the story structure and about the story content. We present
here a brief list of the categories of conjunctions that we have found in
one narrative text, BABAR, and a discussion of what we believe to be their
importance to discourse analysis and discourse understanding.

We omit any discussion of conjoined phrases, e.g., as in the sentence
Here are Cousin Arthur and his friend Zephir, the monkey, since these do not
affect the text as a whole and are strictly sentence-internal conjunctions.
Conjunctions between clauses of the same sentence are included, since the
two clauses could be considered as separate sentences for analyses on the
level of individual propositions.

The most frequent conjunction in BABAR is but. Sentence-initial But
must be distinguished from the use of but to conjoin parallel clauses within
a sentence. The latter usage introduces a clause whose content is contrary
to the expectation that the writer intends to be generated in the reader by

the preceding clause or clauses within the same sentence. Sentence-initial

But makes a similar contrast, but not between the following clause and the

one immediately preceding it, which is in a different sentence, but between

the following clause and some proposition at a higher level of discourse.

This proposition may be a stated one, or one which is merely intended to be

inferred from what is stated somewhere in the preceding discourse.

We begin our analysis of some of the connective functions of but in

BABAR with an instance of sentence-initial But:

> (30-39) (##13) "I need my crown!" says Babar. "I must wear it
> tonight!" "Don't worry," says Celeste. "We'll find that Mustache-
> man. We'll look all over Paris till we do."
>
> (##14) So out they go, looking for the man with Babar's bag.
> "He may be up in the Eiffel Tower," says Celeste. "All visitors
> to Paris go up there."
>
> (##17) Now they are up in the Eiffel Tower. But the man with
> Babar's bag is not.

But here marks the fact that the family expected to find the Mustache-man

in the Eiffel Tower and, contrary to their expectations, did not find him.

The expectation that the man would be there was created by Celeste's state-

ment, "He may be up in the Eiffel Tower," and by the fact that the family

did indeed go up in the Eiffel Tower. (Note that no explicit statement that

the family went up in the Eiffel Tower for the purpose of finding the

Mustache-man is made--this must be inferred by the reader.)

The next occurrence of but follows the previous discourse; i.e., on

pages 17 and 18 we read:

> (38-42) (##17) Now they are up in the Eiffel Tower. But the man
> with Babar's bag is not. "Look at the boats down there!" the
> children shout. "Let's go for a ride."
>
> (##18) Babar is sad, but he goes along.

The interpretation of the clause-conjoining <u>but</u> here requires quite a de-
ductive chain. To fully comprehend this sentence, the reader must assume
that when one is sad, one generally does not feel like doing something
that is enjoyable, or perhaps that when one is sad, one does not feel like
being around other people. Now with this as a cultural presupposition, the
reader must have the world knowledge that taking a boat ride is supposed to
be an enjoyable thing. An instantiation of the general cultural presupposi-
tion, then, leads to an understanding of why Babar's going along for the ride
is contrary to the expectation of the probable behavior of Babar, given the
knowledge that he is sad.

The next occurrences are in the discourse from pages 23 through 27.

(54-70) (##23) The boat does not stop. So Arthur dives off.
"I'll catch him," he cries. "I'll catch that Mustache-man!"
(##24) Arthur climbs out of the water just as fast as he can.
He is all wet. He sees the man with the bag. (##25) Arthur
calls to him. But the man does not hear him. He is walking
away. (##26) Arthur runs after him. He waves. He yells,
"Come back, Mister! You have Babar's crown!" (##27) But
now the man is on a bus. The bus goes down the street. Now
the Mustache-man is gone.

The occurrence of <u>but</u> on p. 25 is simply a contrary-to-expectation conjunc-
tion, which could have joined the clause containing it to the preceding one
in a single sentence. The second occurrence, on page 27, is somewhat more
complicated. <u>But</u> is contrary to expectation, but it is also combined with
<u>now</u> to signal a change in the temporal setting. This <u>but</u> does not crucially
refer to the proposition that the Mustache-man is on the bus, but rather to
an inference that Arthur has failed to get the Mustache-man's attention and
stop him from going away.

The next <u>but</u> is different. On page 28 we find:

> (71-75) Arthur is standing in the sun to dry his clothes.
> Suddenly he sees the whole family. They come running toward
> him. "I'm so glad to see you," he says. "But the red bag--
> it got away."

Here we have a <u>but</u> of <u>contrast</u>. Arthur has stated that he is glad to see

the family. This is a happy event. Now he must tell the family the sad

news, that the bag got away. Notice that we cannot view this <u>but</u> as con-

trary to the expectations of the family, for if that were the case, Arthur

should have been able to use a discourse-initial, "But the red bag got away."

But this sentence would be anomalous in context. Rather, we must see the

conjunction here as contrasting the happy with the sad.

It is also very interesting to look at those contexts where <u>but</u> could

be used and is not, and also those contexts which would be incomprehensible

without the conjunction.

We find one case in BABAR where the conjunction <u>but</u> would fit nicely,

but is not used. This is in the discourse on pages 20 through 23:

> (47-54) (##20) Suddenly Zephir shouts, "Babar, look! Up on
> the bridge! The Mustache-man! He is there with your bag!"
>
> (##21) The children all shout, "Captain, stop! Stop the boat!
> Let us off!"
>
> (##23) The boat does not stop.

<u>But</u> would have fitted quite nicely as the first word of the last sentence

here. The situation is perfect for a contrary-to-expectation conjunction.

We might wonder whether the fact that the conjunction is not present hinders

comprehension.

In the sentence that we have looked at above, <u>Babar is sad, but he goes</u>

<u>along</u>, the conjunction <u>must</u> be included. The discourse would be decidedly

odd if the two clauses were made into separate sentences with no text con-
junction uniting them, i.e.,

> "Look at the boats down there!" the children shout. "Let's
> go for a ride."
>
> Babar is sad. He goes along.

The last two sentences seem entirely unrelated without the conjunction.
Clearly a lot more work needs to be done both on the comprehension of sen-
tence connection with and without conjunctions and on the theoretical lin-
guistic aspects of when conjunctions are required by the grammar and when
they are not. This as an almost entirely uncharted area.

Another interesting kind of text conjunction is one we call a topic
connector conjunction. The sole example in BABAR is on page 52:

> (134-137) Sadly they come up from the subway. Babar says
> nothing. He is very, very sad. And the children are very
> tired.

The last sentence here is the one of interest. The reason it is of such
great interest is that it really does not relate to any of the previous
discourse. The Babar family has been chasing the elusive Mustache-man all
over Paris and has lost him once again in the subway. The topic at the be-
ginning of page 52 is clearly Babar's feelings. The conjunction and of the
final sentence serves to unite the final sentence with the previous state-
ments about Babar by refocussing the topic from Babar's feelings to feelings
of the protagonists in general. Notice that the discourse would sound less
connected without the conjunction, i.e.,

> Sadly they come up from the subway. Babar says nothing.
> He is very, very sad. The children are very tired.

The function of the conjunction and is perhaps one of forcing the cohesion
of topic within the same scene.

106

Now and then, which are traditionally considered to be temporal adverbs, may also be seen as kinds of topic connecting or topic establishing conjunctions. Now, especially, functions throughout BABAR to convey the information that a new scene or new event is about to be introduced. It typically introduces a sentence which describes the accompanying illustration. Without the illustrations, rather long chains of inferences are required to connect half of the sentences beginning with now with the preceding discourse. Some examples:

> (5-8) (##3) Queen Celeste and King Babar pack their crowns. Babar puts his crown in a little red bag.

> (##4) Now the Babar family is on the train. The train is coming into Paris.

The inference must be made that the Babar family got on the train.

> (36-40) (##14) "He may be up in the Eiffel Tower," says Celeste. "All visitors to Paris go up there."

> (##17) Now they are up in the Eiffel Tower. But the man with Babar's bag is not.

Again an inference that the family went up in the Eiffel Tower must be made. Examples like these occur at the beginning of scenes.

The other now's are more difficult to analyze. For example, we see on pages 26 and 27:

> (64-70) Arthur runs after him. He waves. He yells, "Come, back, Mister! You have Babar's crown!

> But now the man is on a bus. The bus goes down the street. Now the Mustache-man is gone.

The first now is a scene-setting conjunction, descriptive of the illustration and requiring an inference that the man got on a bus. The second now is different. It reflects the point of view of the episode-protagonist,

Arthur. Whereas the illustration-descriptive, scene-introduction now refers to the present relative to the reading of the story (that is, it reflects the language a narrator would use if the story were being told or read with the illustrations as cues), this now refers to the present time relative to the event being described.

The conjunction then provides us with some interesting properties as well. It appears that this conjunction can be used to signal a change or break-up in an action sequence. For example,

> (117-123) They get out in front of a market. "I guess we'll have to forget about my crown," sighs Babar. So the children begin to run and play. They race around. They hide behind boxes.
>
> Then they see another man with a small red bag. All the children rush after him.

Notice that without the conjunction then in this discourse, the sequence is incomprehensible, i.e.,

> So the children begin to run and play. They race around. They hide behind boxes. They see another man with a small red bag.

A parallelism is set up here with sentences of similar structure and length, i.e., They race around. They hide behind boxes., and so for the reader to be able to appreciate the importance of the fact that the children see another man with a bag, the author must mark such information as not on a par with the preceding sentences.

Much more work, both theoretical and experimental is necessary before we will be able to describe when or where this kind of marking is required to facilitate comprehension.

c. Anaphoric Devices

A major cohesive device in any text is that of <u>anaphoric reference</u>, or the reference of a noun or noun phrase to some previous mention of that item. The simplest forms of anaphora are those of coreferential nouns or pronouns. For example, in

> (5-6) Queen Celeste and King Babar pack their crowns. Babar puts his crown in a little red bag.

Babar in the second sentence refers anaphorically to the phrase <u>King Babar</u> in the first. <u>His</u> in the phrase <u>his crown</u> also refers to Babar. These anaphoric references are ubiquitous in every kind of text. That is, we typically find pronominal references following close behind their referents.

There are in the texts we examined, however, a fair number of anaphoric references that are not so straightforward. In this section, we will give some representative examples of these and discuss them in terms of the problems such uses of anaphoric devices may pose for comprehension.

1) <u>Reference analysis</u>. One aspect of discourse that can be studied as part of a text analysis program concerns how characters, objects, events, places, etc. are first introduced into the discourse and then later referred to anaphorically in terms of either a definite pronoun or a definite description.[2] Such an analysis can provide the data needed to answer many interesting questions about a text, including the following:

> a. What particular assumptions underlie the use of each definite description in a text and are those assumptions justified? (A definite description may be used either to introduce a character, object, etc. or to refer to it anaphorically. It is a unique description in that in using it, the speaker/writer makes one of two general assumptions. One assumption is that there is one and only one discourse entity which the listener/reader is already aware of that s/he would associate with that description. The other

assumption is that in using it, the listener/reader knows that
there is (or can be) only one object so describable and creates
a unique discourse entity accordingly. Given the first general
assumption, a reference analysis can be used to identify both
the reason that the speaker/writer assumes a definite descrip-
tion will pick out any discourse entity the listener/reader is
aware of and the reason that s/he assumes that it can discrimi-
nate among all those so known. Given the second general assump-
tion, a reference analysis can be used to identify what knowledge
of the world or of English the listener/reader must possess in
order to guarantee such uniqueness.)

b. To what extent are the characters, objects, etc., separable,
on the basis of the descriptions given or derivable from the
text? (Low separability may lead to confusions.)

c. What is the rate at which discourse entities are introduced?
(Too many in rapid succession may be too great a load on pro-
cessing and/or memory.)

d. How much text/narrative intervenes between a discourse enti-
ty's introduction and a subsequent anaphoric reference to it?
(If the gap is large, there may be insufficient content to the
anaphoric expression to find the intended entity.)

e. At any given point in the text, which discourse entities
does the speaker/writer assume the listener/reader is focussed
on such that they can be accessed via the minimal cues of de-
finite pronominal reference? If there are several entities
accessible via the same pronoun, on what basis (contextual
and/or inferential) does the speaker/writer presume the lis-
tener/reader can identify the intended referent? (This is the
pronoun resolution problem discussed at length in the AI and
cognitive psychology literatures (cf. Charniak, 1972; Winograd,
1972; and Chafe, 1974, among others.)

This section is based on preliminary analyses of both Babar Loses His
Crown and the passage entitled "Indian Occupation" (page 38 above) which note
how characters, objects, etc., are introduced contextually and how they are
later referred to anaphorically. In this initial work, we have focussed on
the first question above, since any of a wide range of skills and knowledge
--syntactic, semantic, factual, etc.--may be called upon to justify a partic-
ular definite description.

The analysis is based primarily on research reported on in Nash-Webber (1978a,b). Before describing our method of analysis and its application to these two texts, it would be useful to understand some of the fundamental assumptions guiding our approach to research on reference. The central notion is that of a discourse model. We assume that one objective of discourse is to communicate a model: the speaker/writer has a model of some situation which s/he wishes to communicate to a listener/reader. Thus, the ensuing discourse is, at one level of interaction, an attempt by the speaker to direct the listener in synthesizing a similar model. (In this sense, we are equating "understanding" with "synthesizing an appropriate model.")

Informally, a discourse model may be described as the set of entities "naturally evoked" by a discourse and linked together by the relations they participate in. These are the discourse entities that were mentioned above.[3] In order to understand what we mean by the notion of entities "naturally evoked" by a discourse, consider the following sentence:

1. Each 3rd grade girl brought a brick to Wendy's house.

Then consider each continuation in example 2. In each case, the referents of the definite pronoun (e.g., "she," "it," "they") would be an entity "naturally evoked" by sentence 1.

2a. She certainly was surprised.
     She = Wendy

b. They knew she would be surprised.
    They = the set of 3rd grade girls.

c. She piled them on the front lawn.
    them = the set of bricks, each of which some 3rd grade girl
    brought to Wendy's house

d. She was surprised that they knew where it was.
    it = Wendy's house

e. Needless to say, it surprised her.
    it = the brick-presenting event

Now a speaker/writer is usually not able to communicate at once all the

relevant properties and relations s/he may want to ascribe to any one of

these discourse entities. That task requires multiple acts of reference.

When the speaker/writer wants to refer to an entity in his or her discourse

model, s/he has two ways to do so. One way is with a definite pronoun. In

using a definite pronoun, the speaker/writer assumes 1) that on the basis of

the discourse thus far, a similar entity will be in the listener/reader's

(partially formed) model; and 2) that the listener/reader will be able to

access and identify that entity via the minimal cues of pronominal reference.

The referent of a definite pronoun is thus an entity in the speaker's dis-

course model which s/he presumes to have a counterpart in the listener's

discourse model.

Alternatively, the speaker may refer to an entity in his or her dis-

course model by constructing a description of it in terms of some or all of

its known properties and/or relations (e.g., "a definite pronoun or a def-

inite description").

So while a discourse entity E can be the referent of a definite anaphor

A, we consider A's antecedent to be a unique description of E conveyed to

the listener by the immediately preceding text. The relationship between

the discourse on the one hand and the referents of definite anaphora on the

other is thus a direct one, mediated by the discourse participants' models.

It is our belief that one can formalize, at the sentence level, rules

for deriving unique descriptions of the discourse entities evoked by a text.

A preliminary set of eleven formal rules which are sensitive to such aspects

of a sentence as how each noun phrase is determined, what the relative scope is of each quantifier, and what dependencies exist between noun phrases due to relative clauses can be found in Nash-Webber (1978a,b). As given there, the rules do not take into account tense, modality-belief and deontic-contexts, and certain aspects of negation, all of which can be shown to be necessary factors in forming appropriate unique descriptions. However, in performing our preliminary reference analysis, we have intuitively extended the rules to cover these aspects as well.

As an example of these rules, consider the following one, RW-1, which applies to propositions in which a singular existential quantifier (i.e., a singular indefinite noun phrase) has the widest scope.

(RW-1)   If a proposition $\underline{S}_j$ is of the form

$$(\exists \underline{x} : \underline{A}) \cdot \underline{Px}$$

then it follows that

$$(\exists \underline{x}) \cdot \underline{y} = \underline{\imath z} : \quad \underline{Az} \ \& \ \underline{Pz} \ \& \ \text{evoke} \ \underline{S}_j, \ \underline{z}$$

i.e., informally, if a proposition states that there is a member $\underline{x}$ of class $\underline{A}$ for which $\underline{P}$ is true, then there exists a discourse entity describable as "the $\underline{A}$ which $\underline{Ps}$ which was mentioned (or evoked) by the proposition." (Here $\underline{\imath}$ stands for Russell's definite operator, iota.) Since a unique description can be ascribed to this discourse entity, it can be referred to with a definite anaphor.

There are many places in the Babar story where the application of this rule, RW-1, accounts for both the existence of a new discourse entity and an appropriate unique description for it. A particularly straightforward example is the first sentence of page 20 of the story, which we shall label 20.1 for convenience.

20.1 The boat is going toward a bridge.

As a first approximation--that is finessing the semantics of "going toward"

--sentence 20.1 can be represented as

$(\exists \underline{x}:\text{Bridge})$. Going-toward $\underline{b}_1,\underline{x}$,

where $\underline{b}_1$ is a unique label for the discourse entity referred to anaphorically

with the definite description "the boat."[4] Since this matches the left-hand

side of rule RW-1, it follows that

$(\exists \underline{y})$. $\underline{y} = \underline{i}\underline{z}$: Bridge $\underline{z}$ & Going-toward $\underline{b}_1,\underline{z}$ & evoke $\underline{S}_{20.1},\underline{z}$,

i.e., there exists an individual discourse entity uniquely describable as

"the bridge which was mentioned in Sentence 20.1 which the boat was going

toward." This is the discourse entity referred to anaphorically via the

definite description "the bridge" in the very next sentence.

20.2 Suddenly Zephir shouts, "Babar, look! Up on the bridge."

In our preliminary analysis of the two texts BABAR and INDIAN, our method

of analysis was as follows: we went through the texts sentence by sentence

considering each noun phrase in turn. To each indefinite referential noun

phrase[5] we applied the appropriate rule given in Nash-Webber (1978a), extended

intuitively as mentioned earlier. We assigned the discourse entity that was

evoked a label (e.g., $\underline{e}_1$, $\underline{e}_2$, etc.) and its unique description. For each

definite referential noun phrase, we considered whether it matched the de-

scription of an existing discourse entity and was therefore anaphoric. If

it was not, we created a new discourse entity, labelled it, assigned it its

given definite description, and considered, a propos of question 1, on what

basis that unique description was justified.

For example, consider again sentence 20.1.

20.1 The boat is going toward the bridge.

On the preceding page, the listener/reader has been told

18.2 A boat is ready to take off.

18.3 "Captain, wait for us!" shout the children.

18.4 They climb aboard.

The indefinite noun phrase in sentence 18.2 should have evoked a discourse
entity in the listener/reader's model uniquely describable as "the boat men-
tioned in sentence 18.2 that was ready to take off" (labelled here $\underline{b}_1$). As
for the definite description "the boat" in sentence 20.1, we find that it
matches that discourse entity description (and moreover, no other one). We
therefore take it as an anaphoric reference to that discourse entity under
the assumption that the listener/reader understands sentence 18.2 in accor-
dance with a rule like RW-1.[6] As for the indefinite noun phrase "a bridge,"
we create a new discourse entity and assign it the unique description "the
bridge mentioned in sentence 20.1 that the boat $\underline{b}_1$ is going towards," again
in accordance with RW-1.

There were two sets of data resulting from this preliminary analysis:
1) an annotated text, with each referential noun phrase tagged with the label
of the discourse entity it either evoked or referred to (useful for investi-
gating questions about the rate at which discourse entities are introduced,
the lag between introduction and later anaphoric reference, etc.), and 2) a
list of discourse entities (actually their labels), each with its unique
description (i.e., in accordance with its initial introduction) and any
additional information about it garnered from subsequent anaphoric references

(useful for investigating questions concerning the separability of characters and the justifiability of other definite descriptions).

We will now show how this kind of reference analysis allows us to make some interesting comments about a text that may be involved in its readability. In particular, we will focus on the justifiability of definite descriptions (question 1 above). As stated, the rule RW-1 given earlier is only applicable when an existential noun phrase is understood to have a wide scope over a sentence. (To put this more simply, though less accurately, for the listener/reader to understand an existential as having wide scope is for him or her to assume that the speaker/writer has some particular x in mind which s/he nevertheless cannot refer to as "the x," since it is not unique.) With this in mind, it is clear that the rule RW-1 is not applicable to the second sentence on page 6 of the Babar story:

6.2 The Babar family is waiting for a taxi.

This does not mean that they are waiting for some particular taxi, but rather that they are waiting for any one that comes along. In other words, sentence 6.2 should not evoke a discourse entity uniquely describable as "the taxi mentioned in Sentence 6.2 which the Babar family is waiting for." But now consider the very next sentence of the text.

8.1 The taxi takes them to their hotel.

How might the writer justify this definite reference to "the taxi"? It is not the particular taxi the family is waiting for, since it is not the case that they are waiting for a particular one. However, the author may be assuming that the listener/reader will infer that if the family is waiting for a taxi, then eventually it will be the case that a taxi picks them up.

This latter sentence, "A taxi picks them up," does imply the existence of a discourse entity uniquely describable as "the 'just-mentioned' taxi which picked them = the Babar family up," which is presumably the referent of the definite anaphor "the taxi." Thus, the use of this definite description is motivated by the writer's assumption that the listener/reader both can and will make this plausible, world-knowledge-based inference about the eventual appearance of a taxi. Such inference may demand a high level of sophistication and familiarity with the real world and thus strain the listener/reader's ability to follow the text.

In the "Indian Occupation" passage, there are many examples of definite descriptions which make heavy demands on the listener/reader's knowledge of particular factual events in order to justify and understand them. This may be one reason why the text is so hard to follow. For example, the passage begins with three definite descriptions (here underscored).

The Indians had not heard from the government.

The suit for Alcatraz was still not settled.

Consider first the phrase "the Indians." It is obviously not anaphoric, since it is the first noun phrase in the passage. Therefore, we create a new discourse entity and try to justify the unique description "the Indians." We find that we cannot. Consider next the phrase "the government." It too is not anaphoric, since it does not match in any way the description of the only discourse entity currently around--the one describable as "the Indians." Again we create a new discourse entity and try to justify the unique description. Again we cannot: there is neither a unique government that we as readers know about nor a unique government that we associate with "the

Indians," whoever they are. The same holds true for the definite descrip-
tion "the suit for Alcatraz." In fact, only if the listener/reader already
knows about the Indian occupation of Alcatraz will there be some justifica-
tion for the uniqueness demanded by these definite descriptions. Even then,
it will be an a posteriori justification, since this context is not imparted
until the second sentence. The passage is filled with such "unanchored"
definites, and it is a benefit of this type of reference analysis that it
can help to point out places where world knowledge and inferences based on
it are demanded for understanding.

In the remainder of this section, we shall comment on our preliminary
reference analysis of <u>Babar Loses His Crown</u> vis-à-vis the remaining four
questions posed at the beginning of the section. Questions 2 and 3 concerned
the separability of discourse entities and the rate at which they are intro-
duced. One case of low separability and high rate of introduction involves
the Babar family themselves, who are introduced on the first page of the
text as follows:

    The Babar family is going to Paris.
    Everyone is packing bags.
    Here are the children--Pom, Flora and Alexander. Here are
    Cousin Arthur and his friend Zephir, the monkey.

Nothing here permits the reader/listener to distinguish Pom, Flora, Alexander,
and Cousin Arthur except their names. Moreover, nothing further is said
about any of these four characters individually until page 23.

    The boat does not stop.
    So Arthur dives off.

In preparing our preliminary analysis, we found that we had no recollec-
tion at this point of who Arthur was, although the use of the name with no

further attributes implied he had been introduced earlier in the story. We attribute his evanescence to the number of characters introduced at the same time as he and the lack of distinguishing properties attributed to each one. This is also a comment on the fourth question concerning the lag between a discourse entity's introduction and a subsequent reference to it and the confusions it leads to. Of course, if we had attended to the illustrations there would have been no problem here; all the elephants are pictured almost in every illustration, and the one on page 23 shows "Arthur" diving off "the" boat.

The fifth question concerned focus and the use of pronominal reference, in particular, the case where there is more than one character in focus that is accessible via the same pronoun. Such a case occurs in the Babar story, in a sequence involving Arthur and the man presumed to have taken Babar's crown bag.

24.1 Arthur climbs out of the water just as fast as he can.
24.2 He is all wet.
24.3 He sees the man with the bag.

25.1 Arthur calls to him.
25.2 But the man does not hear him.
25.3 He is walking away.

26.1 Arthur runs after him.
26.2 He waves.

In sentence 24.3, the pronoun "he" is used to refer to Arthur and the definite description is used to refer to the man. In the next sentence (25.1), things switch, and the pronoun "him" is used to refer to the man and the proper name to refer to Arthur. In the next sentence (25.2), things switch again, with the pronoun now referring to Arthur, and the definite description to the man. In the next sentence (25.3), "he" presumably refers to the man,

and two sentences later (26.2), "he" presumably refers to Arthur. The use of the definite description and proper name serve to keep both characters in focus. However, keeping them equally in focus also keeps them equally possible referents for the definite pronoun "he," if context is not sufficient to favor one over the other. The crucial question is whether context is indeed a sufficient cue to the intended referent of "he" in the two sentences--"He is walking away" and "He waves." Further research is obviously necessary, both to refine our methods of reference analysis and to identify some measures of readability associated with reference. One current intuition is that stories beginning "Once upon a time there (was, lived, stood, etc.). . ." may really be easier to understand, at least initially, than ones which do not.

2) <u>Other aspects of anaphora</u>. In this section we take up in more detail two of the questions posed at the beginning of the previous section, and introduce an additional aspect of discourse reference which might be expected to be a source of difficulty in comprehension. We comment first on the subtleties of beginning to solve question 4, how much difference distance between a pronoun and its referent makes. Then we consider certain intricacies of the resolution problem. Finally, we take up the problem of implied referents.

As mentioned in the preceding analysis section, anaphoric reference by pronouns may lead to comprehension difficulty if the pronoun referent is too far removed in time or topic from its original referent, as in this made-up discourse:

> John is one of my best friends. And his sister Suzanna is one of the best tennis players in the country. In fact, she might even make the International Circuit this year if her luck only holds out. She just

> has to win one more tournament in California and then she's a cinch
> to be invited. Anyway, as I was saying, he is one of my best
> friends. . . .

Here, the pronoun "he" in the final sentence is not very comprehensible

after all the information about a different topic, namely Suzanna, has been

presented. Notice that in no way can the pronoun reference be said to be

ambiguous--John is the only possible referent. Still, the time and topic

shift may cause comprehension problems.

In a text like BABAR, we find many, many pronoun references. For the

most part, these references follow closely in time the lexical noun phrases

that they refer to, as in this passage:

> (58-67) Arthur climbs out of the water just as fast as he can. He
> is all wet. He sees the man with the bag. Arthur calls to him.
> He is walking away. Arthur runs after him. He waves. He yells,
> "Come back, Mister. You have Babar's crown!"

With an understanding of the situation, there is little if any difficulty

interpreting the referents of the he's and him's.

But there is one class of pronoun references that one would suspect, on

the basis of distance between references, would cause comprehension diffi-

culties, but which do not in fact do so. This has to do with the references

of the pronoun "they." In this example, which stretches over five pages,

> (128-134) (##48) They all follow him, shouting, "Stop, please,
> Mr. Mustache!" (##50) Too late!" Stuck again! The gates at the
> botton of the stairs snap shut. (##51) "Bring back my crown!"
> shouts Babar. But the man gets on a train, and the train goes
> away. (##52) Sadly, they come up from the subway.

there is quite a bit of intervening material between the first and second

they's. We hypothesize that the reason the pronoun "they" seems so immedi-

ately comprehensible is that "the Babar family" is highly topical, in that

the concept of the Babar family is central to the story and always kept in

mind. Thus, almost any reference to the family in any way, even with a

pronoun, may be expected to be easily understood.

We take up now the subcase of the resolution problem that we may call

contextual redefinition. As an example, let us examine the use of the phrase

the children in BABAR. In this passage from page 2, we see the phrase re-

ferring specifically to three individuals, Pom, Flora, and Alexander.

(3-4) Here are the children--Pom, Flora and Alexander. Here
are Cousin Arthur and his friend Zephir, the monkey.

But it has a different reference in the following passage:

(117-125) They get out in front of a market. "I guess we'll have
to forget about my crown," sighs Babar. So the children begin to
run and play. They race around. They hide behind boxes. Then they
see another man with a small red bag. All the children rush after
him. Arthur knocks over a box of apples. Zephir knocks over a box
of fish.

Here children, at least by the second occurrence of the word, seems to refer

to all five of the younger protagonists. This point may be argued here,

but not in another example,

(138-141) Celeste says, "We'll put the children to bed in the hotel.
Then we'll leave them and go to the opera." Back in their hotel
room, they say goodnight to the children. The three littlest ones
are already fast asleep.

Now exactly who is tired or who has been running and playing in the market

is not terribly crucial to the story here. Nevertheless, the contextual re-

definition of particular lexical items is seen in these examples, and uses

of the same lexical phrase for different references is a possibility in gen-

eral and can lead to ambiguities or confusions of reference. The same prob-

lem of contextual redefinition of reference is, of course, always present

for pronouns, usually to an even greater extent than for lexical noun phrases.

Finally, we take up the problem of anaphoric reference to <u>implied ref-</u>
<u>erents</u>. This is exemplified by instances where there is an anaphoric con-
nection to some object that is not overtly given in the text, but rather
must be inferred by the reader, as in this passage from BABAR:

> (11-13) Now they are off the train, with all their bags. The
> Babar family is waiting for a taxi. The taxi takes them to
> their hotel.

Notice that the Babar family is not waiting for any specific taxi. Rather,
they are waiting for taxi transportation. In the sentence <u>The taxi takes</u>
<u>them to their hotel</u>, the taxi referred to must be some taxi which finally
picked them up, and the phrase <u>the taxi</u> must be short for "the taxi that
picked up the family." No prior reference is made to this taxi and no men-
tion is made overtly in the story that there was a taxi that picked up the
Babar family. The reference to <u>taxi</u> is, then, a reference to an implied
referent.

Although the question of the proper logical and psychological analysis
of implied reference is a difficult one, with frame analysis offering one
promising possibility, our suspicion is that references to implied referents
will not typically create comprehension difficulties, except possibly where
knowledge of the frame referred to is imperfect.

C. <u>Text-Level Inferences</u>

In this section we discuss the analysis of inferences that have to be
made in comprehension of text which is intended as connected discourse. Our
analyses are not intended to provide an exhaustive listing of all possible
inferences that could or must be made from the use of individual sentences.
Rather, we have concentrated on enumerating inferences that must be made in

order to understand the (intended) relation between (or among) the sentences of a text; in other words, inferences that the author must have implicitly intended to be made, as opposed to ones which are merely possible, given the meanings of the lexical items and the rules of syntax.

As an illustration, let us consider a couple of cases. The book <u>Babar</u> <u>Loses His Crown</u> begins with the following sentence:

    1.   The Babar family is going to Paris.

Among the inferences that it is possible to make from this sentence are the following:

    1a.   There is a Babar family.
    1b.   There are at least two individuals related by blood or marriage who are going to Paris.
    1c.   Babar is their last name.
    1d.   Paris is a place. (It could be a city, a country, a planet, a neighborhood, a restaurant, or a department store, <u>apud alia</u>.)
    1e.   They are not in/at/on Paris now.

But inference (1b) follows as an automatic consequence of understanding the meaning of the word <u>family</u>, (1d) and (1e) follow from the meanings of <u>going</u> and <u>to</u>, and (1a) follows from linguistic conventions on the use of the word <u>the</u>. Inference (1c), which happens to be incorrect (Babar is the father's first name), follows from culture-specific linguistic conventions about referring to families. But what all of these inferences have in common is that they are inferences that can or must be made of the sentence <u>in isolation</u>: the same inference could be expected to be made if the sentence had been uttered as the initiation of a conversation. In other words, these inferences are independent of the <u>use</u> of the sentence in a connected piece of discourse. Since ability to make these inferences follows automatically from knowledge of the language plus the ability to reason, they do not have the

124

direct kind of bearing on reading competence that inferences about the re-
lations among sentences have.

To contrast these types of inferences, let us examine another case.
These two sentences are also from BABAR.

93. But Babar can't eat.
94. He is thinking about his crown.

The following inferences may be made about the relation between sentences
(93) and (94):

94a. Babar is thinking about his crown.
94b. The reason Babar can't eat is that he is thinking about his crown.
94c. Thinking about his crown makes it impossible for Babar to eat.

Inference (94a) is not of so much interest here as it is an inference of the
reference of the pronoun he, and depends on knowledge of conditions on the
use of pronouns discussed in Section IV B 3 of this report. Inference (94b)
is presumably the one the author intended the reader to make: he presumably
intended sentence (94) as an explanation for the state reported in sentence
(93). What about inference (94c)? While (94c) might be taken as being a
necessary premise, along with (94a), for concluding (94b), which follows not
from (94a) and (94c) taken as premises, but from the fact that the author has
provided (93) and (94) in that order, and from the assumption that he is
adhering to a principle of cooperative discourse which provides that unless
warning is given to the contrary, a sentence will be assumed to be relevant
to the point of previous discourse (cf. Grice, 1975). From this fact and
this assumption, the reader may infer that the relevance of (94) to (93) is
that (94) is intended as an explanation for (93). Note that this is not a
necessary inference like (94a); the reader might be wrong about the author;

(94) might not have been intended to be relevant to (93). Or (94) might have been intended to be relevant only insofar as it described a state of Babar simultaneous with (93). We conclude, however, that (94b) is an inference that should be made for this text (BABAR) to be understood--that the author probably did intend the two sentence (93) and (94) to be taken as being related by (94b).

Thus our enumeration of text-level inferences concentrates on inferences relating sentences to each other. The enumerations are probably not complete, however; indeterminacy and dependence on the intentions of an unavailable individual, the author, of the sort just described prevents this from being a feasible goal. We have, nonetheless, attempted to provide accounts of the inferences that must be made in order to understand the sequences of sentences as coherent texts representing the product of some purposeful activity.

V. Illustration and Layout Analysis

The importance of the role of linguistic properties of written materials for readability is obvious. Less obvious, but possibly quite important, is the role of non-linguistic aspects of the text, ranging from the size and color of the pages and choice of type face to relations between the language of the text and its visual properties. In this section we explore two problems that belong in this domain: illustrations and their relation to the text, and the relation between line-layout and certain syntactic and semantic properties.

In the first section, we present a discussion of the analysis of illustrations and the relations of their content to the understanding of the text.

In the second section we present discussion of the relation of line-ends to syntactic structure and to the structure of direct quotation.

A. Illustrations

1. Relevance of Illustrations to Comprehension

Illustrations in children's books are not merely decorative enticements to read. In scientific and social studies texts, and even, perhaps surprisingly, in many narratives, they represent an integral part of the content. It is typically the case with preprimers and basal readers for the first few grades that the texts are incomprehensible without reference to the accompanying illustrations, and even when the texts are self-contained, illustrations may be sufficiently redundant that a child could become proficient enough at interpreting illustrations to lose interest in improving his ability to decode written language. For this reason, we consider the relation between a text and accompanying illustrations to be an important object of analysis.

Section 2 describes the procedures used in analysis of the redundancy of illustrations to text, and some issues that will have to be resolved or accommodated before meaningful comparisons of texts can be attempted.

2. Discussion of the Calculation of the Portion of a Narrative Retrievable from Accompanying Illustrations

a. Introduction

Two investigators calculated the proportion of the text retrievable in the illustrated narrative (Babar Loses His Crown) according to the procedure specified vaguely as follows:

The Procedure

1. For each illustration, provide content statements which tell the story told by the accompanying text. In other words, translate the text into a sequence of statements, which recounts the story in the same order as the text, but with no direct discourse (except where essential to the plot), and no rhetorical devices. Statements should describe one action or setting each.

2. For each statement, indicate what portion of the content of that statement is retrievable from the accompanying illustration.

3. Average the percent retrievable for all statements to obtain a measure of the amount of the text retrievable from the accompanying illustrations.

The vagueness of this specification of the procedure is attested by the fact that the initial assessments by the investigators differed by more than 30%: the research assistant calculated that the illustrations told 62% of the story; the principal investigator calculated that they told roughly 40%. After discussing the rationale for the assessments with each other, correcting inconsistencies and over- and under-evaluations that could be agreed upon, the assessments were 52% and 48%, respectively.

b. Problems of Analysis

Several types of problems arose in the attempt to assess the amount of a narrative story that was retrievable from accompanying illustrations, and ultimately to arrive at the specifiable procedure for making this assessment. We describe these in some detail.

1) The translation of text into content statements. The first problem that arises, both logically and procedurally, is how to translate the text into content statements. It had been agreed that the content statement should be a natural narrative unit, but there were no criteria for determining how much text a statement could encompass, or for constraining the form of

statements beyond the rule of thumb that the number of statements would

generally correspond to the number of finite verbs in the text, though they

need not always, as in the following examples from other texts:

    1. One statement, more than one finite verb

        Text: Debbie had a dream. It was a
        bad dream. It was about a tiger.

        Statement: Debbie had a bad dream about a tiger.

    2. One finite verb, more than one statement

        Text: Mog was nice--but not very clever.

        Statements: Mog was nice. But Mog was not very clever.

Direct discourse was a problem in its own right. Our attempt was to

attend to the perlocutionary effects of quoted speech as they would appear

in a recounting of the narrative, thus:

    3. Direct discourse

        Text: "Pardon me," said the fiddler to the milkman.
        "I am a stranger in town. Perhaps you could direct
        me to a place where I might have breakfast."

        Statements: The fiddler introduced himself to the
        milkman. He asked where he could have breakfast.

Notice that statements cannot usefully be held to "propositions" in the

logician's usage. It does not help to decompose statements like <u>He asked</u>

<u>where he could have breakfast</u> into component propositions. Likewise, if

one follows the Generative Semantics program of semantic analysis, every pre-

dicative element (including quantifiers, conjunctions, negatives, and ad-

verbs, as well as attributive adjectives and nouns) is the nucleus of some

semantic proposition, but it appears that an analysis at this level would

fragment information too much to be useful for our purposes.

Given the limitations that had been established, however vaguely, and the shared exemplars, it did not appear that BABAR would present too many problems of translation into statements, as there are hardly any sentences with more than one surface clause (13 out of 191, counting infinitive phrases as clauses). However, even so syntactically impoverished a text as BABAR turned out to present problems.

To begin with the simplest, consider the case of modifiers that are not syntactically obligatory, e.g., the underlined phrases in sentences 11 and 17:

11. Now they are off the train with all their bags.

17. In their hotel room, Celeste opens all their bags.

One investigator treated with all their bags as a separate statement (all their bags were with them), as it turns out to be important to the story, and in their hotel room as part of a statement, as it was unimportant to the narrative where the unpacking took place. The other investigator treated with all their bags as part of a statement, apparently because a retelling of the narrative with that content as a separate sentence accords more importance to this information than the original text does. This investigator, on the other hand, treated in their hotel room as a separate statement ([they go] to their hotel room; bracketed material refers to content that must be inferred) since this could be recounted separately without affecting the development of the plot or sounding bizarre. The investigators were unable to agree on which criteria were more valid, as each has both advantages and disadvantages as summarized below.

a. Importance of the content to the narrative: if it is important, count it as a separate statement. This seems plausible, but

there are two hitches. First, it requires foreknowlege (which a
first-time reader does not have): you have to know what happens
later in order to know what is important at the beginning. On the
other hand, if comprehension works in such a way that the reader's
model of what the important things are is continually being revised,
then he has hindsight by the time he is finished, which accomplishes
the same thing. Second, making importance to the story critical
would result in a content analysis where some predicative elements
are the nuclei of content statements but many others have their
(surface) nominal or attributive roles; some phrases in text sen-
tences would be semantically decomposed into several statements
while others would be parts of undecomposed statements.

b. Coherence of the sequence of content statements as a text:
do not make it a separate statement if it would make the sequence
of content statements an incoherent text. This "transderivational"
criterion may turn out to be unempirical, but its purpose was to
provide for an analysis which looked like it had a chance of cor-
responding to a skeleton narrative from which the story could be
elaborated, or reconstructed, combining statements in complex
sentences.

c. Rhetorical effect of decomposition: do not make it a separate
statement if it results in emphases different from those in the
original text. Again a "transderivational" criterion, this has
less likelihood of being useful, as almost any meddling with a
text is bound to have some rhetorical effect.

2) Totally irrelevant material. Some textual material is totally

irrelevant to the story being told, for example, Last of all in sentence 18.

18. Last of all, she opens the little red one.

The question arises whether this would be considered part of the content of

the narrative. While this issue makes a difference here, since it affects

the proportion of content retrievable from illustrations, its relevance is

not limited to investigation of inferences from illustrations, but will

recur wherever it is attempted to objectify, quantify, or otherwise analyze

content. Some books have more irrelevant content than others. The under-

lined phrases in the following passages from Howard Garls' Uncle Wiggly and

His Friends (1939, Platt & Munk) were all judged to be entirely irrelevant

to the stories they occur in.

> One day Uncle Wiggly Longears started out for a ride in his
> automobile. It had a turnip steering wheel that he could
> nibble on when he was hungry. . . . Pretty soon he came to
> a place where there was a little shop, made from corncobs.
>
> Then the muskrat lady, who was also camping on the island,
> began to cook breakfast. The kind circus elephant [never
> before mentioned, note] got a pail of water in his trunk, at
> the well Uncle Wiggly had dug.

In fact, the whole breakfast episode in the second passage is entirely

irrelevant to the story it occurs in.

3) Proper names. A third problem, which is limited to research on

illustrations, is how to evaluate text which includes the names of individ-

uals who figure in the narrative, particularly the material which introduces

these individuals. That is, when you have a text like sentences 3 and 4

from BABAR, and an illustration showing three smallish elephants (two dressed

in shorts, one in a dress), one larger elephant dressed in a (French) sailor

suit, and a monkey, in shorts and a beret, how should the information content

of the illustration with respect to this text be counted?

> 3. Here are the children--Pom, Flora and Alexander.

> 4. Here are Cousin Arthur and his friend Zephir, the monkey.

All that a "picture-reader" can tell at this point is that some juvenile

elephants and a monkey are involved (along with two adult elephants) in the

story. This was counted as equivalent to about half of the content of sen-

tence 3 and one-quarter of sentence 4. The underlined portion of the

following statements recounting sentences 3 and 4 was considered to be repre-

sented in the illustrations.

> There are some (elephant) children. They are Pom, Flora, and
> Alexander. There is a cousin and a monkey. The cousin's name
> is Arthur. The monkey is Cousin Arthur's friend. His name is
> Zephir.

<center>or</center>

<u>There are (elephant) children</u>--Pom, Flora and Alexander. <u>There</u>
<u>is</u> a cousin (Arthur) and his friend Zephir, <u>the monkey</u>.

It must be noted that even a text-reader cannot tell which elephant has

which name, not even which is Cousin Arthur, the monkey's friend. Later, if

the reader pays close attention to the pictures, he or she will be able to

infer which names go with certain elephants, but even after hundreds of

readings, it is possible to identify with certainty only Flora and Arthur

without accompanying text, in addition to the obvious Babar and Celeste. In

any case, the names of individuals is something a picture-reader would never

be able to infer from illustrations. What does this imply for the evaluation

of the retrievability of later text which predicates various actions of these

individuals, referring to them by name? Not much, it was concluded, since it

seemed most reasonable to assume that as far as the <u>story</u> retrievable from

the text was concerned, picture-readers would distinguish the characters by

means of identifying-expressions which referred to the illustrations, on the

order of "the elephant with the sailor suit" or "one of the little-boy ele-

phants," and they would be able to tell an equivalent story about the same

individuals. They would just refer to them differently from the way the text

does.

4) <u>Text which requires inference for interpretation</u>. A number of the

difficulties which arose involved inferences from text or illustrations (or

perhaps, speaking more correctly, chains of inferences, since interpretation

of straightforward text and simple illustration always involves making some

inferences). To begin with, there are cases in which information must be

inferred from the text in making the translation to content statement. But

it is not clear how much is inference and how much is just understanding

sentence meaning. For example, consider the illustration on pages 10-11.

> Celeste (profile view) is seated in front of an open red
> suitcase, grasping a flute in her trunk. Her eyebrow (not
> shown in other illustrations; maybe it is a forehead wrinkle)
> is raised (in surprise?). The other five elephants stand
> opposite, their trunks extended (in interest?), Babar's at
> nearly 90 degrees.

The accompanying text includes sentences 19 through 26.

19. "Look!" she cries.

20. "What is this? 21. A flute!

22. Babar! 23. This is not your bag."

24. "My crown. 25. It's lost!" cries Babar.

26. "My crown is gone!"

The point of sentences 19-21 is that Celeste has discovered a flute (in the

red bag), and that of sentences 24-26, that Babar is very upset because his

crown is not there. In principle, much more of the inferable content can

be retrieved from suitable illustrations than from such fragmentary exclama-

tory direct discourse as is found here. In this case, however, only the ex-

istence of the flute was counted as retrievable (and not its discovery, as

Celeste's eyebrow is neither very salient nor unambiguous). Babar's dismay

could have been pictured, but it was judged not to be.

5) Knowledge of the world required for the fullest interpretation of

illustration. A thornier problem involves cases where there is a reasonable

question as to whether a picture-reader can be expected to be able to make

certain inferences from an illustration. The answer to this question depends

on how much real-world knowledge can be imputed to the picture-reader, as-

suming that he has mastered the complex strategies for interpretation of

abstract representations of real-world objects. (Informal observations suggest that 15-month-old infants can do this.) In BABAR, this problem arose with reference to the taxicab and hotel scenes. For instance, on pages 6-7 there is the following sentence:

12. The Babar family is waiting for a taxi.

The illustration shows the Babar family standing next to a group of humans. At the extreme left is the front one-third of an automobile with a "knob" on top which has the word TAXI written on it. Is it plausible to expect a picture-reader (who cannot read TAXI) to recognize that this auto is a taxi, and infer that the Babar family is waiting for it? [It is not clear that even that inference is invited--they could be waiting for the light to change. It may be only our knowledge that frequently when one takes a train trip (cf. pages 4-5), one takes a taxi to a more "specific" destination that makes this inference not unreasonable. But this is treated elsewhere.) The point is, what does an illiterate have to know about taxis to recognize this object as a taxi? Is the half-inch by three-eighths inch "knob" with letters on it enough? Does the short-billed hat on the head of the driver help enough? Or is it part of the meaning of taxi that taxis have "knobs" on top?

The problems in the next scene are probably more evident. Sentences 13 and 14 read:

13. The taxi takes them to their hotel.

14. Celeste and the children walk inside.

Disregarding the taxi (part of the right front fender and grill are shown), the illustration shows Celeste entering a stone or concrete edifice through one of two doors in arched doorways festooned with Neo-Gothic or Baroque or

Roman heads and iron grill work. Part of a wrought-iron balcony is shown above the doors. In other words, it is a stereotypical European hotel. Can a picture-reader be expected to recognize it as a hotel? Even if he or she knows what a hotel is, it does not bear much resemblance to an American hotel (outside of New York, anyway).

6) <u>Illustrations that are fully interpretable only by inferences relating them to preceding illustrations.</u> The final identified problems involve illustrations that cannot be interpreted without reference to previous illustrations. In the illustrations just described, if one infers that the 1/25 of an auto shown on page 8 is the same as the 1/3 of a taxi shown on page 6, then it is reasonable to infer that it brought the elephants to the location shown on pages 8-9. But if one does not connect the illustrations like this, it is just 1/25 of an auto. Continuing, pages 10-11 show the elephants, a chair, a hassock, and an open red suitcase. If the reader interprets pages 8-9 as saying that the elephants enter a hotel, and connects it to this illustration, he may be able to infer the content of the initial adverbial phrase in sentence 17.

17. In their hotel room, Celeste opens all the bags.

But without pages 8-9, the illustration on pages 10-11 says nothing <u>directly</u> about the location of the scene. The chair, hassock, and open suitcase invite an inference of a room, perhaps even a room where they are scheduled to spend the night, but no background is shown at all--no walls, no windows.

Pages 48-51 exemplify the opposite situation--an illustration that can be interpreted correctly (with respect to the accompanying text) only by taking into consideration a subsequent illustration. On page 48 we have an

illustration of the Babar family running after a man who is running down some stairs which are flanked by a sign which says METRO, accompanied by the text:

126. Now the whole Babar family chases the man with the bag.

127. Down the stairs of the subway.

128. They all follow him, shouting, "Stop, please, Mr. Mustache!"

Nothing in this illustration (and precious little in the text, taken at its face value) indicates that the Babar family descend the stairs. The illustration on pages 50-51 shows the heads of Babar, Arthur, Zephir, and Celeste peeping over a gate which encloses them in a sort of tunnel which is hung with signs reading CORRESPONDANCE and SORTIE. The man (notice that this is a reference to the previous illustration) is shown stepping through a sliding door (more specific, real-world knowledge); an adjacent window shows a seated couple; presumably this is a subway train (more real-world subcultural knowledge). The point is, if one can infer that this is a subway, can one infer, with respect to the previous page, that the Babar family did follow the man down the stairs, which were stairs to the subway? Making the inference is not made any easier by the fact that no stairs are shown in the illustration on pages 50-51.

7) Illustrated content that has no apparent relevance to the narrative until a number of frames after its occurrence. The most difficult problem involves information that is present and clear, but whose relevance does not become clear to the picture-reader until many pages later, if at all. Take the case of the flute and the Mustache-man (the crux of the story). Page 10 shows Celeste holding up a flute with a look of mild surprise in her eye. No

flute recurs until page 60, where Babar is shown holding one/it, with a
smile on his face, confronting the/a mustachioed man, who is smiling and
holding a crown. Pages 62-63 show Babar and Celeste in a box at a Wagnerian
opera. The orchestra pit is shown, and the flute-player has a mustache.
The flute is about one and one-quarter inches by one-sixteenth of an inch;
the mustache three-sixteenths of an inch by one-sixteenth, so neither is
very salient. Would or could a picture-reader, on encountering the illus-
trations on pages 60-63, remember back to page 10, and deduce in essence
the plot of the whole story: that the two red bags (on page 6!) got switched,
and that the Babar family had been running all over the city trying to find
the Mustache-man who they thought would have Babar's crown (most recently
seen on page 3). Could he or she figure from pages 60-63 that it was not
the family's flute, and therefore not their bag, and that therefore someone
else must have had their bag?

A case which appears only slightly less difficult perceptually is that
of pages 10-13, where the text indicates that Babar has lost his crown.
Could a person infer from Celeste's surprise (granting that she would inter-
pret that much) on page 10, and Babar's sadness on pages 12-13 (the flute is
not shown; the red bag is closed; otherwise the scene is the same as pages
10-11), that Babar was sad because his crown was not in the bag? To do so,
he or she would have to remember Babar packing his crown, which was shown, but
not very saliently, on page 3.

Likewise, the content of sentence 87 on pages 34-35

87. "He is not the Mustache-man at all!"

would be retrievable from the illustrations of the Babar family surrounding

a clean-shaven man, if the picture-reader knew that the Babar family had been looking for a mustachioed man.[7] He might have inferred this from the sequence of illustrations on pages 24-27. These show (pages 24-25) Arthur climbing out of the canal/river (the Seine actually), and walking in the direction of a man with a mustache and a red bag who appears to be about to board a bus. The Mustache-man and the bus make up about 1/8 of the illustration. Pages 26-27 show Arthur running after a bus which is carrying the/a mustachioed man. The red bag is not shown. Is Arthur's running after this man enough to make the chase the central issue in the succeeding pages for the picture-reader? (The/a mustachioed man, who first appears on pages 4-5, appears again before this scene on page 21, where he is one of three background figures, taking up, himself, perhaps 1/200 of the illustration.) It is not clear from the illustration (though perhaps it is supposed to be) that the elephants who are in a boat see him and his red suitcase, and that this prompts Arthur to jump off the boat to get to him (pages 22-23).

c. Additional Kinds of Text-Illustration Relations

1) Text which functions as description of accompanying illustration. A number of sentences in BABAR, and other Babar books, can be considered grammatically correct only if construed as being descriptions of the accompanying illustration of the sort that might be provided by someone interpreting for a small child a story in a picture-book with no text. Examples include sentences 3, 4, 7, 11, 38, 68, 71, 110, and 126:

3. Here are the children--Pom, Flora and Alexander. (Pages 2-3)

4. Here are Cousin Arthur and his friend Zephir, the monkey. (Pages 2-3)

7. Now the Babar family is on the train. (Pages 4-5)

11. Now they are off the train. (Pages 6-7)

38. Now they are up in the Eiffel Tower. (Pages 16-17)

68. But now the man is on a bus. (Pages 26-27)

71. Arthur is standing in the sun to dry out his clothes. (Pages 28-29)

110. But now a red light. (Pages 42-43)

126. Now the whole Babar family chases the man with the bag. (Pages 48-49)

These often are in the present progressive, like sentence 71, or contain certain deictic elements which can be interpreted only as referring to the illustrations. In all of these cases, the illustration depicts the situation described in the now-sentence. If it did not, such narrative now-sentences could only be grammatically construed as equivalent to then-statements if the narrative is taken to be in the "historical present" which is used when the events described are particularly real and vivid to the narrator, and the sequence of events is very fast-paced. This construal of these sentences can be ruled out here with some certainty, as the sequence of events unfolds rather slowly, and there is no evidence that the narrator is at all personally involved in any of the events recounted. His presence is inferable only from the sentences which describe the illustrations, and the few evaluative remarks found in such sentences as 115 and 160.

115. Poor Babar!

160. It turns out to be a great night after all.

In each case, the now also marks a change of scene in the story, sometimes

with a considerable gap in time, and a gap in the text which the reader has
to fill in by inference; for example, on pages 2-3, sentences 2 and 7:

    2. Everyone is packing bags.

    7. Now the Babar family is on the train.

Nothing was said about how they got on the train, or what they did between
packing and getting on the train. Likewise, with reference to sentences 36-
38 on pages 14-17, it must be inferred that the family in fact decided to go
look for the Mustache-man in the Eiffel Tower.

    36. "He may be in the Eiffel Tower," says Celeste.

    37. "All visitors to Paris go up there."

    38. Now they are up in the Eiffel Tower.

    2) <u>Supplementary information in illustrations</u>. There are no major
parts of the text which are incomprehensible without the illustrations. The
illustrations seem to show a general picture of what is happening, each two-
page illustration depicting a new scene in the action recounted by the story.
Thus, the main function of the illustrations, besides a decorative one, seems
to be to give an actual visual representation of the individuals, objects,
and events referred to in the story. Thus, they might help the child in
understanding such new notions as "Eiffel Tower," "hotel," and "sidewalk
restaurant," which might not be familiar to him or her. The whole situation of
chasing the Mustache-man down the stairs to the subway and being stopped by
the gates at the bottom of the stairs might be almost incomprehensible to a
child not familiar with the French Metro system. The illustrations here
supplement the text by showing parts of these events. In this case neither
the illustration alone nor the text alone will suffice to provide enough

information for an untravelled six- or seven-year-old to fully comprehend the situation described; only if a child relates them to each other will he or she be able to make inferences from the pictures to the text, and vice-versa, to make sense of what is happening.

The same kind of text-illustration relation is found on the last two pages of the story (pages 62-63). On pages 60-61, Babar recovers his crown after bumping into the Mustache-man, who gets his flute back. Pages 62-63 contain the following text:

160. It turns out to be a great night after all.

161. The crown is on the head of the King . . . and the flute is under the Mustache-man's mustache.

The reader would have a difficult time inferring from the text alone that Babar and Celeste did get to the opera after all, since the only indication of this in the text is the mention of the flute being under the Mustache-man's mustache. Inferences are required along the following lines: If the flute is under his mustache, perhaps it is in his mouth, since the mouth is under the mustache. If it is in his mouth, he is probably playing it. Perhaps he is in the orchestra at the opera. Since the location of the crown with respect to the King (Babar) is mentioned in the same sentence, perhaps the King is at the opera too. This last is perhaps still an improbable inference, but enhanced nevertheless by the fact that Babar and Celeste were on their way to the opera. In principle, of course, Babar could have decided to have a "great night" at a nightclub to celebrate the recovery of his crown. The illustration suggests the correct inferences by showing Babar with his crown sitting in a box at the opera and a mustachioed flautist in the orchestra pit with the rest of the orchestra, and this helps the reader make a little more sense of the last sentence of the story.

142

d. Further research

    There are perhaps more problems in content and inference analysis of
text and accompanying illustration which further work will identify. Certain
relatively naturalistic experiments could perhaps be devised to answer some
of the questions raised above, but it is not clear how much could be general-
ized from the answers.

    If it is possible to devise procedures which will permit the comparison
of illustrated texts with respect to the redundancy of accompanying illustra-
tions, then experiments can be designed to show what properties of illus-
trated texts aid in the acquisition of reading ability, and under what con-
ditions, and what properties retard or discourage it.

B. Layout

    1. Introduction

    Descriptions of the layout of a text should include mention of at least
the following properties which may contribute to the readability of the text:

    1. Presence of illustrations

    2. Relative portions of the page occupied by text, illustrations,
        and "white space"

    3. Margins (right justified or not)

    4. Size of type

    5. Some measure of the number of lines, words per page

    6. Paragraphing

    7. Front matter (title page, background, etc.)

    8. Back matter (index, comprehension questions, etc.)

We discuss here only one of the most interesting and potentially impor-
tant aspects of layout: the effect of line breaks on readability. It is
likely that the interruption at line ends can cause an interruption in lan-
guage processing. Experimentation is needed to determine whether the exact
placement of line ends (their coincidence, or lack of it, with crucial points
in syntactic and semantic structure, for instance) is an important factor in
readability. Most likely any effects would be most pronounced in beginning
readers.

Here we explore the correlations between line ends and syntactic struc-
ture, and between line ends and certain aspects of the structure of direct
quotations. Anyone who has read much modern poetry has noticed that the
selection of line ends can be exploited for effects that vary from poetry
to gimmickry. It is clear that certain combinations of line end and syn-
tactic properties can have a jarring effect on the mature reader. It may be
that this effect is the symptom of some hang-up in syntactic processing
caused by the line end interruption, which could obstruct normal processing
and comprehension in the immature reader. These judgments are intuitive and
impressionistic; experimental work is clearly needed to determine the impor-
tance of this factor.

An analogous question arises in regard to the relation between line end
interruptions and other kinds of linguistic properties. We have taken as a
case in point the relation between direct quotation and line ends; it is
conceivable that the processing task of keeping track of direct quotations,
their contents, and associated speakers, is made more difficult by the inter-
ruption of line ends.

2. Syntactic Description of Line Ends

Probably because of short line length, many sentences were broken at line ends. In the Analysis of Babar (Tech. Rep. No. 000), broken sentences are listed with surface-structure parsings, end-of-line being indicated by "+". The depth in the tree representing the surface syntactic structure of the sentence at which the break occurred is indicated. We suspect that the relation between line breaks and various syntactic properties is significant for beginning readers, in that the line break is likely to cause an interruption in syntactic (and other) processing. The analyst recorded impressionistic judgments of the jarring effect of line breaks, on a scale of 1 to 7. This rough guess of level of difficulty does not entirely correlate with depth. This is a matter that deserves detailed linguistic and experimental investigation.

3. Quotations and Line Ends

A matter analogous to the relation between syntax and line ends is the role of line ends and other interruptions in the presentation of direct quotation; the reader's task is simultaneously to keep track of who is speaking and to process and comprehend what is said. This complex task can possibly become more difficult when a single monologue by the same speaker is broken by punctuation, syntax (e.g., Quote-niching, as in sentence 144), or line ends. Also, readers could easily be confused by unannounced speaker changes, as in the transition from Babar to Celeste in sentences 31 and 32 of BABAR. As it turns out, in BABAR all such speaker-to-speaker transitions involve either pre-announcement of new speaker (as in sentences 50-51) or separation of the two speakers by a page break. The question of possible effects on beginning readers deserves experimental examination.

References

Abelson, R. P., Aronson, E., McGuire, W. J., Newcomb, T. M., Rosenberg, M. J.,

    & Tannenbaum, P. H. Theories of cognitive consistency: A sourcebook.

    Chicago: Rand McNally, 1968.

Abrahamson, A. A. Experimental analysis of the semantics of movement. In

    D. A. Norman & D. E. Rumelhart (Eds.), Explorations in cognition. San

    Francisco: Freeman, 1975.

Bartlett, R. C. Remembering: A study in experimental and social psychology.

    Cambridge: Cambridge University Press, 1932.

Bowerman, M. The acquisition of word meanings: An investigation of some

    current conflicts. Paper presented at the Third International Child

    Language Symposium, London, September 1975.

Bruce, B. C. What makes a good story? Language Arts, 1978, 55, 460-466.

Bruce, B. C. Analysis of interacting plans as a guide to the understanding

    of story structure (Tech. Rep. No. 130). Urbana: University of Illinois,

    Center for the Study of Reading, June 1979. (ERIC Document Reproduction

    Service No. ED 174 951)

Bruce, B. C., & Newman, D. Interacting plans (Tech. Rep. No. 88). Urbana:

    University of Illinois, Center for the Study of Reading, June 1978.

    (ERIC Document Reproduction Service No. ED 157 038)

Chafe, W. L. Meaning and the structure of language. Chicago: University of

    Chicago Press, 1970.

Chafe, W. L. Language and consciousness. Language, 1974, 50, 11-133.

Chafe, W. L. Givenness, contrastiveness, definiteness, subjects, topics, and

    point of view. In C. Li (Ed.), Subject and topic. New York: Academic

    Press, 1976.

Charniak, E. Towards a model of children's story comprehension (Tech. Rep. No. 266). Cambridge, Mass.: M.I.T. Artificial Intelligence Laboratory, 1972.

Clark, E. V. What's in a word: On the child's acquisition of semantics in his first language. In T. E. Moore (Ed.), Cognitive development and the acquisition of language. New York: Academic Press, 1973.

Cohen, P. R., & Perrault, C. R. Preliminaries for a computer model of conversation. In Proceedings of the CSCSI National Conference. Vancouver, Canada, August 1976.

Daneš, F. Functional sentence perspective and the organization of the text. In F. Daneš (Ed.), Papers on functional sentence perspective. Prague: Academia, 1974. Pp. 106-128.

Gentner, D. Evidence for the psychological reality of semantic components: The verbs of possession. In D. A. Norman & D. E. Rumelhart (Eds.), Explorations in cognition. San Francisco: Freeman, 1975.

Gentner, D. On relational meaning: The acquisition of verb meaning. Child Development, 1978, 48, 988-998.

Grice, H. P. Logic and conversation. In P. Cole & J. L. Morgan (Eds.), Syntax and semantics (Vol. 3): Speech acts. New York: Academic Press, 1975.

Gundel, J. Role of topic and comment in linguistic theory. Bloomington: Indiana University Linguistics Club, 1977.

Halliday, M. A. K. Notes on transitivity and theme in English, part II. Journal of Linguistics, 1967, 3, 199-247.

147

Hermon, G. On the discourse structure of direct quotation (Tech. Rep. No. 143). Urbana: University of Illinois, Center for the Study of Reading, September 1979. (ERIC Document Reproduction Service No. ED 177 495)

Hunt, K. W. Grammatical structures written at three grade levels (NCTE Research Report No. 3). Urbana, Ill.: National Council of Teachers of English, 1965.

Kantor, R. Introducing referents and degrees of activatedness. Paper presented at N-WAVE VI, Georgetown University, Washington, D.C., 1977.

Karttunen, L. Discourse referents. In J. McCawley (Ed.), Syntax and semantics (Vol. 7): Notes from the linguistic underground. New York: Academic Press, 1976.

Keenan, E. O., & Schieffelin, B. Topic as a discourse notion: A study of topic in the conversations of children and adults. In C. Li (Ed.), Subject and topic. New York: Academic Press, 1976.

Kintsch, W. The representation of meaning in memory. Hillsdale, N.J.: Erlbaum, 1974.

Li, C., & Thompson, S. Subject and topic: A new typology of language. In C. Li (Ed.), Subject and topic. New York: Academic Press, 1976.

Mandler, J. M., & Johnson, N. S. A structural analysis of stories and their recall: From "once upon a time" to "happily ever after." Cognitive Psychology, 1977, 9, 111-151.

Meyer, B. J. F. The organization of prose and its effects on memory. Amsterdam: North Holland, 1975.

Miller, G. A. English verbs of motion: A case study in semantics and lexical memory. In A. W. Melton & E. Martin (Eds.), Coding processes in human memory. Washington, D.C.: Winston, 1972.

Miller, G. A. Images and models, similes and metaphors. In A. Ortony (Ed.), Metaphor and thought. Cambridge: Cambridge University Press, 1979.

Miller, G. A., & Johnson-Laird, P. N. Language and perception. Cambridge, Mass.: Harvard University Press, 1976.

Morgan, J. L., & Sellner, M. B. Discourse and linguistic theory. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension. Hillsdale, N.J.: Erlbaum, in press.

Munro, A. Linguistic theory and the LNR structural representation. In D. A. Norman & D. E. Rumelhart (Eds.), Explorations in cognition. San Francisco: Freeman, 1975.

Nash-Webber, B. L. Inference in an approach to discourse anaphora (Tech. Rep. No. 77). Urbana: University of Illinois, Center for the Study of Reading, January 1978. (a) (ERIC Document Reproduction Service No. ED 150 552)

Nash-Webber, B. L. A formal approach to discourse anaphora. Unpublished doctoral dissertation, Harvard University, 1978. (b)

Propp, V. I. Morphology of the folk-tale (Publication 10). Bloomington: Indiana University, Research Center in Anthropology, Folklore and Linguistics, 1958.

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

Rumelhart, D. E., & Levin, J. A. A language comprehension system. In D. A. Norman & D. E. Rumelhart (Eds.), Explorations in cognition. San Francisco: Freeman, 1975.

Rumelhart, D. E., & Norman, D. A. The computer implementation. In D. A. Norman & D. E. Rumelhart (Eds.), Explorations in cognition. San Francisco: Freeman, 1975.

Rumelhart, D. E., & Ortony, A. The representation of knowledge in memory (CHIP Report No. 55). La Jolla: University of California, Center for Human Information Processing, 1976.

Sacerdoti, E. D. Planning in a hierarchy of abstraction spaces. In Proceedings of the Third International Conference on Artificial Intelligence. Stanford, California, August 1973.

Sacerdoti, E. D. The nonlinear nature of plans (Tech. Rep. No. 101). Menlo Park, Calif.: Stanford Research Institute, Artificial Intelligence Group, January 1975.

Schmidt, C. F., & Sridharan, N. S. The representation of plans: Rule of consistency for plan recognition. In Proceedings of AISB Summer Conference. Edinburgh, Scotland, July 1976.

Schulz, C. He's your dog, Charlie Brown. New York: New American Library, 1974.

Stein, N. L. The comprehension and appreciation of stories: A developmental analysis. In S. Madeja (Ed.), The arts and cognition (Vol. 2). St. Louis, Mo.: Cemrel, 1978.

Stein, N. L., & Glenn, C. G. An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), Multidisciplinary approaches to discourse comprehension. Hillsdale, N.J.: Ablex, 1977.

Stein, N. L., & Nezworski, T. The effects of organization and instructional set on story memory. Discourse Processes, 1978, 1, 177-193.

Sussman, G. A computer model of skill acquisition. New York: American

   Elsevier, 1975.

Tate, A. Interacting goals and their use. In Proceedings of the Fourth

   International Joint Conference on Artificial Intelligence. Tbillisi,

   Georgia, U.S.S.R., September 1975.

Winograd, T. Understanding natural language. New York: Academic Press,

   1972.

Footnotes

[1] Detailed analyses of some aspects of BABAR (<u>Babar Loses His Crown</u>) and DESERT ("The Wonderful Desert") are given in Reading Center Technical Reports 169 and 170.

[2] For convenience, we shall often use the term "discourse entity" to refer to any such individual, object, event, etc. evoked by the text.

[3] They play a somewhat similar role to Karttunen's "discourse referents" (1976). Our alternate terminology rests on a desire to keep "referents" a separate technical term.

[4] See Nash-Webber (1978a) for a full explanation of the notation used above. As using it here unmotivated and unexplained may lead to more confusion than it clarifies, in the remainder of this section, we will appeal to the reader's intuitions about the results of applying RW-1.

[5] Most noun phrases are referential. Non-referential noun phrases include predicate nominatives and appositives.

[6] We are not claiming that he or she consciously applies rule RW-1, an obvious idealization, to some formal representation of that sentence, just that his or her understanding of the sentence has the same effect as RW-1.

[7] A similar problem arises with sentence 39:

39. The man with the bag is not there.

Only, in the case of 39, there is nothing in any illustration to indicate that they are looking for a man with a bag. Of course, this is consistent with the illustration of pages 16-17, but so are an infinite number of other content statements, many no more or less likely than this, e.g., Celeste's crown is not there, President de Gaulle is not there, etc.

Appendix I: Hansel and Gretel

Close to a large forest there lived a woodcutter with his wife and his two children. The boy was called Hansel and the girl Gretel. They were always very poor and had very little to live on. And at one time when there was famine in the land, he could no longer procure daily bread.

One night when he lay in bed worrying over his troubles, he sighed and said to his wife, "What is to become of us? How are we to feed our poor children when we have nothing for ourselves?"

"I'll tell you what, husband," answered the woman. "Tomorrow morning we will take the children out quite early into the thickest part of the forest. We will light a fire and give each of them a piece of bread. Then we will go to our work and leave them alone. They won't be able to find their way back, and so we shall be rid of them."

"Nay, wife," said the man, "we won't do that. I could never find it in my heart to leave my children alone in the forest. The wild animals would soon tear them to pieces."

"What a fool you are!" she said. "Then we must all four die of hunger. You may as well plane the boards for our coffins at once."

She gave him no peace till he consented. "But I grieve over the poor children all the same," said the man. The two children could not go to sleep for hunger either, and they heard what their stepmother said to their father.

Gretel wept bitterly and said, "All is over with us now."

"Be quiet, Gretel," said Hansel. "Don't cry! I will find some way out of it."

When the old people had gone to sleep, he got up, put on his little coat, opened the door, and slipped out. The moon was shining brightly and the white pebbles round the house shone like newly minted coins. Hansel stooped down and put as many into his pockets as they would hold.

Then he went back to Gretel and said, "Take comfort, little sister, and go to sleep. God won't forsake us." And then he went to bed again.

At daybreak, before the sun had risen, the woman came and said, "Get up, you lazybones! We are going into the forest to fetch wood."

Then she gave them each a piece of bread and said, "Here is something for your dinner, but don't eat it before then, for you'll get no more."

Gretel put the bread under her apron, for Hansel had the stones in his pockets. Then they all started for the forest.

When they had gone a little way, Hansel stopped and looked back at the cottage, and he did the same thing again and again.

His father said, "Hansel, what are you stopping to look back at? Take care and put your best foot foremost."

"Oh, father," said Hansel, "I am looking at my white cat. It is sitting on the roof, wanting to say good-by to me."

"Little fool, that's no cat! It's the morning sun shining on the chimney," said the mother.

But Hansel had not been looking at the cat. He had been dropping a pebble on the ground each time he stopped.

When they reached the middle of the forest, their father said, "Now, children, pick up some wood. I want to make a fire to warm you."

154

Hansel and Gretel gathered the twigs together and soon made a huge pile. Then the pile was lighted, and when it blazed up the woman said, "Now lie down by the fire and rest yourselves while we go and cut wood. When we have finished we will come back to fetch you."

Hansel and Gretel sat by the fire, and when dinnertime came they each ate their little bit of bread, and they thought their father was quite near because they could hear the sound of an ax. It was no ax, however, but a branch which the man had tied to a dead tree, and which blew backwards and forwards against it. They sat there so long a time that they got tired. Then their eyes began to close and they were soon fast asleep.

When they woke it was dark night. Gretel began to cry, "How shall we ever get out of the wood?"

But Hansel comforted her and said, "Wait a little while till the moon rises, and then we will soon find our way."

When the full moon rose, Hansel took his little sister's hand and they walked on, guided by the pebbles, which glittered like newly coined money. They walked the whole night, and at daybreak they found themselves back at their father's cottage.

They knocked at the door, and when the woman opened it and saw Hansel and Gretel she said, "You bad children, why did you sleep so long in the wood? We thought you did not mean to come back any more."

But their father was delighted, for it had gone to his heart to leave them behind alone.

155

Appendix II: The Dog and the Cock

From The Winston Readers: First Reader by S. G. Firman, copyright © 1918, Holt, Rinehart, and Winston.

Once a dog and a cock went into the woods.

Soon it grew dark.

The cock said, "Let us stay here all night. I will sleep in this tree-top. You can sleep in the hollow trunk."

"Very well," said the dog.

So the dog and the cock went to sleep.

In the morning the cock began to crow, "Cock-a-doodle-do! Cock-a-doodle-do!"

Mr. Fox heard him crow.

He said, "That is a cock crowing. He must be lost in the woods. I will eat him for my breakfast."

Soon Mr. Fox saw the cock in the tree-top.

He said to himself, "Ha! ha! Ha! ha! What a fine breakfast I shall have! I must make him come down from the tree. Ha! ha! Ha! ha!"

So he said to the cock, "What a fine cock you are! How well you sing! Will you come to my house for breakfast?"

The cock said, "Yes, thank you, I will come if my friend may come, too."

"Oh yes," said the fox. "I will ask your friend. Where is he?"

The cock said, "My friend is in this hollow tree. He is asleep. You must wake him."

Mr. Fox said to himself, "Ha! ha! I shall have two cocks for my breakfast!"

So he put his head into the hollow tree.

Then he said, "Will you come to my house for breakfast?"

Out jumped the dog and caught Mr. Fox by the nose.

CENTER FOR THE STUDY OF READING

READING EDUCATION REPORTS

No. 1: Durkin, D. *Comprehension Instruction—Where are You?*, October 1977. (ERIC Document Reproduction Service No. ED 146 566, 14p., PC-$1.82, MF-$.83)

No. 2: Asher, S. R. *Sex Differences in Reading Achievement*, October 1977. (ERIC Document Reproduction Service No. ED 145 567, 30p., PC-$3.32, MF-$.83)

No. 3: Adams, M. J., Anderson, R. C., & Durkin, D. *Beginning Reading: Theory and Practice*, November 1977. (ERIC Document Reproduction Service No. ED 151 722, 15p., PC-$1.82, MF-$.83)

No. 4: Jenkins, J. R., & Pany, D. *Teaching Reading Comprehension in the Middle Grades*, January 1978. (ERIC Document Reproduction Service No. ED 151 756, 36p., PC-$3.32, MF-$.83)

No. 5: Bruce, B. *What Makes a Good Story?*, June 1978. (ERIC Document Reproduction Service No. ED 158 222, 16p., PC-$1.82, MF-$.83)

No. 6: Anderson, T. H. *Another Look at the Self-Questioning Study Technique*, September 1978. (ERIC Document Reproduction Service No. ED 163 441, 19p., PC-$1.82, MF-$.83)

No. 7: Pearson, P. D., & Kamil, M. L. *Basic Processes and Instructional Practices in Teaching Reading*, December 1978. (ERIC Document Reproduction Service No. ED 165 118, 29p., PC-$3.32, MF-$.83)

No. 8: Collins, A., & Haviland, S. E. *Children's Reading Problems*, June 1979. (ERIC Document Reproduction Service No. ED 172 188, 19p., PC-$1.82, MF-$.83)

No. 9: Schallert, D. L., & Kleiman, G. M. *Some Reasons Why Teachers are Easier to Understand than Textbooks*, June 1979. (ERIC Document Reproduction Service No. ED 172 189, 17p., PC-$1.82, MF-$.83)

No. 10: Baker, L. *Do I Understand or Do I not Understand: That is the Question*, July 1979. (ERIC Document Reproduction Service No. ED 174 948, 27p., PC-$3.32, MF-$.83)

No. 11: Anderson, R. C., & Freebody, P. *Vocabulary Knowledge and Reading*, August 1979.

No. 12: Joag-dev, C., & Steffensen, M. S. *Studies of the Bicultural Reader: Implications for Teachers and Librarians*, January 1980.

No. 13: Adams, M., & Bruce, B. *Background Knowledge and Reading Comprehension*, January 1980.

No. 14: Rubin, A. *Making Stories, Making Sense*, January 1980.

No. 15: Tierney, R. J., & LaZansky, J. *The Rights and Responsibilities of Readers and Writers: A Contractual Agreement*, January 1980.

CENTER FOR THE STUDY OF READING

TECHNICAL REPORTS

No. 1: Halff, H. M. *Graphical Evaluation of Hierarchical Clustering Schemes,* October 1975. (ERIC Document Reproduction Service No. ED 134 926, 11p., PC-$1.82, MF-$.83)

No. 2: Spiro, R. J. *Inferential Reconstruction in Memory for Connected Discourse,* October 1975. (ERIC Document Reproduction Service No. ED 136 187, 81p., PC-$6.32, MF-$.83)

No. 3: Goetz, E. T. *Sentences in Lists and in Connected Discourse,* November 1975. (ERIC Document Reproduction Service No. ED 134 927, 75p., PC-$4.82, MF-$.83)

No. 4: Alessi, S. M., Anderson, T. H., & Biddle, W. B. *Hardware and Software Considerations in Computer Based Course Management,* November 1975. (ERIC Document Reproduction Service No. ED 134 928, 21p., PC-$1.82, MF-$.83)

No. 5: Schallert, D. L. *Improving Memory for Prose: The Relationship between Depth of Processing and Context,* November 1975. (ERIC Document Reproduction Service No. ED 134 929, 37p., PC-$3.32, MF-$.83)

No. 6: Anderson, R. C., Goetz, E. T., Pichert, J. W., & Halff, H. M. *Two Faces of the Conceptual Peg Hypothesis,* January 1976. (ERIC Document Reproduction Service No. ED 134 930, 29p., PC-$3.32, MF-$.83)

No. 7: Ortony, A. *Names, Descriptions, and Pragmatics,* February 1976. (ERIC Document Reproduction Service No. ED 134 931, 25p., PC-$1.82, MF-$.83)

No. 8: Mason, J. M. *Questioning the Notion of Independent Processing Stages in Reading,* February 1976. (*Journal of Educational Psychology,* 1977, *69,* 288-297)

No. 9: Siegel, M. A. *Teacher Behaviors and Curriculum Packages: Implications for Research and Teacher Education,* April 1976. (ERIC Document Reproduction Service No. ED 134 932, 42p., PC-$3.32, MF-$.83)

No. 10: Anderson, R. C., Pichert, J. W., Goetz, E. T., Schallert, D. L., Stevens, K. C., & Trollip, S. R. *Instantiation of General Terms,* March 1976. (ERIC Document Reproduction Service No. ED 134 933, 30p., PC-$3.32, MF-$.83)

No. 11: Armbruster, B. B. *Learning Principles from Prose: A Cognitive Approach Based on Schema Theory,* July 1976. (ERIC Document Reproduction Service No. ED 134 934, 48p., PC-$3.32, MF-$.83)

No. 12: Anderson, R. C., Reynolds, R. E., Schallert, D. L., & Goetz, E. T. *Frameworks for Comprehending Discourse,* July 1976. (ERIC Document Reproduction Service No. ED 134 935, 33p., PC-$3.32, MF-$.83)

No. 13: Rubin, A. D., Bruce, B. C., & Brown, J. S. *A Process-Oriented Language for Describing Aspects of Reading Comprehension,* November 1976. (ERIC Document Reproduction Service No. ED 136 188, 41p., PC-$3.32, MF-$.83)

No. 14: Pichert, J. W., & Anderson, R. C. *Taking Different Perspectives on a Story,* November 1976. (ERIC Document Reproduction Service No. ED 134 936, 30p., PC-$3.32, MF-$.83)

No. 15: Schwartz, R. M. *Strategic Processes in Beginning Reading,* November 1976. (ERIC Document Reproduction Service No. ED 134 937, 19p., PC-$1.82, MF-$.83)

No. 16: Jenkins, J. R., & Pany, D. *Curriculum Biases in Reading Achievement Tests,* November 1976. (ERIC Document Reproduction Service No. ED 134 938, 24p., PC-$1.82, MF-$.83)

No. 17: Asher, S. R., Hymel, S., & Wigfield, A. *Children's Comprehension of High- and Low-Interest Material and a Comparison of Two Cloze Scoring Methods,* November 1976. (ERIC Document Reproduction Service No. ED 134 939, 32p., PC-$3.32, MF-$.83)

No. 18: Brown, A. L., Smiley, S. S., Day, J. D., Townsend, M. A. R., & Lawton, S. C. *Intrusion of a Thematic Idea in Children's Comprehension and Retention of Stories,* December 1976. (ERIC Document Reproduction Service No. ED 136 189, 39p., PC-$3.32, MF-$.83)

No. 19: Kleiman, G. M. *The Prelinguistic Cognitive Basis of Children's Communicative Intentions,* February 1977. (ERIC Document Reproduction Service No. ED 134 940, 51p., PC-$4.82, MF-$.83)

No. 20: Kleiman, G. M. *The Effect of Previous Context on Reading Individual Words,* February 1977. (ERIC Document Reproduction Service No. ED 134 941, 76p., PC-$6.32, MF-$.83)

No. 21: Kane, J. H., & Anderson, R. C. *Depth of Processing and Interference Effects in the Learning and Remembering of Sentences,* February 1977. (ERIC Document Reproduction Service No. ED 134 942, 29p., PC-$3.32, MF-$.83)

No. 22: Brown, A. L., & Campione, J. C. *Memory Strategies in Learning: Training Children to Study Strategically,* March 1977. (ERIC Document Reproduction Service No. ED 136 234, 54p., PC-$4.82, MF-$.83)

No. 23: Smiley, S. S., Oakley, D. D., Worthen, D., Campione, J. C., & Brown, A. L. *Recall of Thematically Relevant Material by Adolescent Good and Poor Readers as a Function of Written Versus Oral Presentation,* March 1977. (ERIC Document Reproduction Service No. ED 136 235, 23p., PC-$1.82, MF$-.83)

No. 24: Anderson, R. C., Spiro, R. J., & Anderson, M. C. *Schemata as Scaffolding for the Representation of Information in Connected Discourse,* March 1977. (ERIC Document Reproduction Service No. ED 136 236, 18p., PC-$1.82, MF-$.83)

No. 25: Pany, D., & Jenkins, J. R. *Learning Word Meanings: A Comparison of Instructional Procedures and Effects on Measures of Reading Comprehension with Learning Disabled Students,* March 1977. (ERIC Document Reproduction Service No. ED 136 237, 34p., PC-$3.32, MF-$.83)

No. 26: Armbruster, B. B., Stevens, R. J., & Rosenshine, B. *Analyzing Content Coverage and Emphasis: A Study of Three Curricula and Two Tests,* March 1977. (ERIC Document Reproduction Service No. ED 136 238, 22p., PC-$1.82, MF-$.83)

No. 27: Ortony, A., Reynolds, R. E., & Arter, J. A. *Metaphor: Theoretical and Empirical Research,* March 1977. (ERIC Document Reproduction Service No. ED 137 752, 63p., PC-$4.82, MF-$.83)

No. 28: Ortony, A. *Remembering and Understanding Jabberwocky and Small-Talk,* March 1977. (ERIC Document Reproduction Service No. ED 137 753, 36p., PC-$3.32, MF-$.83)

No. 29: Schallert, D. L., Kleiman, G. M., & Rubin, A. D. *Analysis of Differences between Oral and Written Language,* April 1977. (ERIC Document Reproduction Service No. ED 144 038, 33p., PC-$3.32, MF-$.83)

No. 30: Goetz, E. T., & Osborn, J. *Procedures for Sampling Texts and Tasks in Kindergarten through Eighth Grade,* April 1977. (ERIC Document Reproduction Service No. ED 146 565, 80p., PC-$6.32, MF-$.83)

No. 31: Nash-Webber, B. *Anaphora: A Cross-Disciplinary Survey,* April 1977. (ERIC Document Reproduction Service No. ED 144 039, 43p., PC-$3.32, MF-$.83)

No. 32: Adams, M. J., & Collins, A. *A Schema-Theoretic View of Reading Comprehension,* April 1977. (ERIC Document Reproduction Service No. ED 142 971, 49p., PC-$3.32, MF-$.83)

No. 33: Huggins, A. W. F. *Syntactic Aspects of Reading Comprehension,* April 1977. (ERIC Document Reproduction Service No. ED 142 972, 68p., PC-$4.82, MF-$.83)

No. 34: Bruce, B. C. *Plans and Social Actions,* April 1977. (ERIC Document Reproduction Service No. ED 149 328, 45p., PC-$3.32, MF-$.83)

No. 35: Rubin, A. D. *Comprehension Processes in Oral and Written Language,* April 1977. (ERIC Document Reproduction Service No. ED 150 550, 61p., PC-$4.82, MF-$.83)

No. 36: Nash-Webber, B., & Reiter, R. *Anaphora and Logical Form: On Formal Meaning Representation for Natural Language,* April 1977. (ERIC Document Reproduction Service No. ED 142 973, 42p., PC-$3.32, MF-$.83)

No. 37: Adams, M. J. *Failures to Comprehend and Levels of Processing in Reading,* April 1977. (ERIC Document Reproduction Service No. ED 145 410, 51p., PC-$4.82, MF-$.83)

No. 38: Woods, W. A. *Multiple Theory Formation in High-Level Perception,* April 1977. (ERIC Document Reproduction Service No. ED 144 020, 58p., PC-$4.82, MF-$.83)

No. 40: Collins, A., Brown, J. S., & Larkin, K. M. *Inference in Text Understanding,* December 1977. (ERIC Document Reproduction Service No. ED 150 547, 48p., PC-$3.32, MF-$.83)

No. 41: Anderson, R. C., & Pichert, J. W. *Recall of Previously Unrecallable Information Following a Shift in Perspective,* April 1977. (ERIC Document Reproduction Service No. ED 142 974, 37p., PC-$3.32, MF-$.83)

No. 42: Mason, J., Osborn, J., & Rosenshine, B. *A Consideration of Skill Hierarchy Approaches to the Teaching of Reading,* December 1977. (ERIC Document Reproduction Service No. ED 150 549, 176p., PC-$12.32, MF-$.83)

No. 43: Collins, A., Brown, A. L., Morgan, J. L., & Brewer, W. F. *The Analysis of Reading Tasks and Texts,* April 1977. (ERIC Document Reproduction Service No. ED 145 404, 96p., PC-$6.32, MF-$.83)

No. 44: McClure, E. *Aspects of Code-Switching in the Discourse of Bilingual Mexican-American Children,* April 1977. (ERIC Document Reproduction Service No. ED 142 975, 38p., PC-$3.32, MF-$.83)

No. 45: Schwartz, R. M. *Relation of Context Utilization and Orthographic Automaticity in Word Identification,* May 1977. (ERIC Document Reproduction Service No. ED 137 762, 27p., PC-$3.32, MF-$.83)

No. 46: Anderson, R. C., Stevens, K. C., Shifrin, Z., & Osborn, J. *Instantiation of Word Meanings in Children,* May 1977. (ERIC Document Reproduction Service No. ED 142 976, 22p., PC-$1.82, MF-$.83)

No. 47: Brown, A. L. *Knowing When, Where, and How to Remember: A Problem of Metacognition,* June 1977. (ERIC Document Reproduction Service No. ED 146 562, 152p., PC-$10.82, MF-$.83)

No. 48: Brown, A. L., & DeLoache, J. S. *Skills, Plans, and Self-Regulation,* July 1977. (ERIC Document Reproduction Service No. ED 144 040, 66p., PC-$4.82, MF-$.83)

No. 49: Goetz, E. T. *Inferences in the Comprehension of and Memory for Text,* July 1977. (ERIC Document Reproduction Service No. ED 150 548, 97p., PC-$6.32, MF-$.83)

No. 50: Anderson, R. C. *Schema-Directed Processes in Language Comprehension,* July 1977. (ERIC Document Reproduction Service No. ED 142 977, 33p., PC-$3.32, MF-$.83)

No. 51: Brown, A. L. *Theories of Memory and the Problems of Development: Activity, Growth, and Knowledge,* July 1977. (ERIC Document Reproduction Service No. ED 144 041, 59p., PC-$4.82, MF-$.83)

No. 52: Morgan, J. L. *Two Types of Convention in Indirect Speech Acts,* July 1977. (ERIC Document Reproduction Service No. ED 145 405, 40p., PC-$3.32, MF-$.83)

No. 53: Brown, A. L., Smiley, S. S., & Lawton, S. C. *The Effects of Experience on the Selection of Suitable Retrieval Cues for Studying from Prose Passages,* July 1977. (ERIC Document Reproduction Service No. ED 144 042, 30p., PC-$3.32, MF-$.83)

No. 54: Fleisher, L. S., & Jenkins, J. R. *Effects of Contextualized and Decontextualized Practice Conditions on Word Recognition,* July 1977. (ERIC Document Reproduction Service No. ED 144 043, 37p., PC-$3.32, MF-$.83)

No. 55: Jenkins, J. R., & Larson, K. *Evaluating Error Correction Procedures for Oral Reading,* June 1978. (ERIC Document Reproduction Service No. ED 158 224, 34p., PC-$3.32, MF-$.83)

No. 56: Anderson, T. H., Standiford, S. N., & Alessi, S. M. *Computer Assisted Problem Solving in an Introductory Statistics Course,* August 1977. (ERIC Document Reproduction Service No. ED 146 563, 26p., PC-$3.32, MF-$.83)

No. 57: Barnitz, J. *Interrelationship of Orthography and Phonological Structure in Learning to Read,* August 1977. (ERIC Document Reproduction Service No. ED 150 546, 62p., PC-$4.82, MF-$.83)

No. 58: Mason, J. M. *The Role of Strategy in Reading in the Mentally Retarded,* September 1977. (ERIC Document Reproduction Service No. ED 145 406, 28p., PC-$3.32, MF-$.83)

No. 59: Mason, J. M. *Reading Readiness: A Definition and Skills Hierarchy from Preschoolers' Developing Conceptions of Print,* September 1977. (ERIC Document Reproduction Service No. ED 145 403, 57p., PC-$4.82, MF-$.83)

No. 60: Spiro, R. J., & Esposito, J. J. *Superficial Processing of Explicit Inferences in Text,* December 1977. (ERIC Document Reproduction Service No. ED 150 545, 27p., PC-$3.32, MF-$.83)

No. 65: Brewer, W. F. *Memory for the Pragmatic Implications of Sentences,* October 1977. (ERIC Document Reproduction Service No. ED 146 564, 27p., PC-$3.32, MF-$.83)

No. 66: Brown, A. L., & Smiley, S. S. *The Development of Strategies for Study Prose Passages,* October 1977. (ERIC Document Reproduction Service No. ED 145 371, 59p., PC-$4.82, MF-$.83)

No. 68: Stein, N. L., & Nezworski, T. *The Effects of Organization and Instructional Set on Story Memory,* January 1978. (ERIC Document Reproduction Service No. ED 149 327, 41p., PC-$3.32, MF-$.83)

No. 69: Stein, N. L. *How Children Understand Stories: A Developmental Analysis,* March 1978. (ERIC Document Reproduction Service No. ED 153 205, 68p., PC-$4.82, MF-$.83)

No. 76: Thieman, T. J., & Brown, A. L. *The Effects of Semantic and Formal Similarity on Recognition Memory for Sentences in Children,* November 1977. (ERIC Document Reproduction Service No. ED 150 551, 26p., PC-$3.32, MF-$.83)

No. 77: Nash-Webber, B. L. *Inferences in an Approach to Discourse Anaphora,* January 1978. (ERIC Document Reproduction Service No. ED 150 552, 30p., PC-$3.32, MF-$.83)

No. 78: Gentner, D. *On Relational Meaning: The Acquisition of Verb Meaning,* December 1977. (ERIC Document Reproduction Service No. ED 149 325, 46p., PC-$3.32, MF-$.83)

No. 79: Royer, J. M. *Theories of Learning Transfer,* January 1978. (ERIC Document Reproduction Service No. ED 149 326, 55p., PC-$4.82, MF-$.83)

No. 80: Arter, J. A., & Jenkins, J. R. *Differential Diagnosis-Prescriptive Teaching: A Critical Appraisal,* January 1978. (ERIC Document Reproduction Service No. ED 150 578, 104p., PC-$7.82, MF-$.83)

No. 81: Shoben, E. J. *Choosing a Model of Sentence Picture Comparisons: A Reply to Catlin and Jones,* February 1978. (ERIC Document Reproduction Service No. ED 150 577, 30p., PC-$3.32, MF-$.83)

No. 82: Steffensen, M. S. *Bereiter and Engelmann Reconsidered: The Evidence from Children Acquiring Black English Vernacular,* March 1978. (ERIC Document Reproduction Service No. ED 153 204, 31p., PC-$3.32, MF-$.83)

No. 83: Reynolds, R. E., Standiford, S. N., & Anderson, R. C. *Distribution of Reading Time When Questions are Asked about a Restricted Category of Text Information,* April 1978. (ERIC Document Reproduction Service No. ED 153 206, 34p., PC-$3.32, MF-$.83)

No. 84: Baker, L. *Processing Temporal Relationships in Simple Stories: Effects of Input Sequence,* April 1978. (ERIC Document Reproduction Service No. ED 157 016, 54p., PC-$4.82, MF-$.83)

No. 85: Mason, J. M., Knisely, E., & Kendall, J. *Effects of Polysemous Words on Sentence Comprehension,* May 1978. (ERIC Document Reproduction Service No. ED 157 015, 34p., PC-$3.32, MF-$.83)

No. 86: Anderson, T. H., Wardrop, J. L., Hively W., Muller, K. E., Anderson, R. I., Hastings, C. N., & Fredericksen, J. *Development and Trial of a Model for Developing Domain Referenced Tests of Reading Comprehension,* May 1978. (ERIC Document Reproduction Service No. ED 157 036, 69p., PC-$4.82, MF-$.83)

No. 87: Andre, M. E. D. A., & Anderson, T. H. *The Development and Evaluation of a Self-Questioning Study Technique,* June 1978. (ERIC Document Reproduction Service No. ED 157 037, 37p., PC-$3.32, MF-$.83)

No. 88: Bruce, B. C., & Newman, D. *Interacting Plans,* June 1978. (ERIC Document Reproduction Service No. ED 157 038, 100p., PC-$6.32, MF-$.83)

No. 89: Bruce, B. C., Collins, A., Rubin, A. D., & Gentner, D. *A Cognitive Science Approach to Writing,* June 1978. (ERIC Document Reproduction Service No. ED 157 039, 57p., PC-$4.82, MF-$.83)

No. 90: Asher, S. R. *Referential Communication,* June 1978. (ERIC Document Reproduction Service No. ED 159 597, 71p., PC-$4.82, MF-$.83)

No. 91: Royer, J. M., & Cunningham, D. J. *On the Theory and Measurement of Reading Comprehension,* June 1978. (ERIC Document Reproduction Service No. ED 157 040, 63p., PC-$4.82, MF-$.83)

No. 92: Mason, J. M., Kendall, J. R. *Facilitating Reading Comprehension Through Text Structure Manipulation,* June 1978. (ERIC Document Reproduction Service No. ED 157 041, 36p., PC-$3.32, MF-$.83)

No. 93: Ortony, A., Schallert, D. L., Reynolds, R. E., & Antos, S. J. *Interpreting Metaphors and Idioms: Some Effects of Context on Comprehension,* July 1978. (ERIC Document Reproduction Service No. ED 157 042, 41p., PC-$3.32, MF-$.83)

No. 94: Brown, A. L., Campione, J. C., & Barclay, C. R. *Training Self-Checking Routines for Estimating Test Readiness: Generalization from List Learning to Prose Recall,* July 1978. (ERIC Document Reproduction Service No. ED 158 226, 41p., PC-$3.32, MF-$.83)

No. 95: Reichman, R. *Conversational Coherency,* July 1978. (ERIC Document Reproduction Service No. ED 159 658, 86p., PC-$6.32, MF-$.83)

No. 96: Wigfield, A., & Asher, S. R. *Age Differences in Children's Referential Communication Performance: An Investigation of Task Effects,* July 1978. (ERIC Document Reproduction Service No. ED 159 659, 31p., PC-$3.32, MF-$.83)

No. 97: Steffensen, M. S., Jogdeo, C., & Anderson, R. C. *A Cross-Cultural Perspective on Reading Comprehension,* July 1978. (ERIC Document Reproduction Service No. ED 159 660, 41p., PC-$3.32, MF-$.83)

No. 98: Green, G. M. *Discourse Functions of Inversion Construction,* July 1978. (ERIC Document Reproduction Service No. ED 160 998, 42p., PC-$3.32, MF-$.83)

No. 99: Asher, S. R. *Influence of Topic Interest on Black Children and White Children's Reading Comprehension,* July 1978. (ERIC Document Reproduction Service No. ED 159 661, 35p., PC-$3.32, MF-$.83)

No. 100: Jenkins, J. R., Pany, D., & Schreck, J. *Vocabulary and Reading Comprehension: Instructional Effects,* August 1978. (ERIC Document Reproduction Service No. ED 160 999, 50p., PC-$3.32, MF-$.83)

No. 101: Shoben, E. J., Rips, L. J., & Smith, E. E. *Issues in Semantic Memory: A Response to Glass and Holyoak,* August 1978. (ERIC Document Reproduction Service No. ED 159 662, 85p., PC-$6.32, MF-$.83)

No. 102: Baker, L., & Stein, N. L. *The Development of Prose Comprehension Skills,* September 1978. (ERIC Document Reproduction Service No. ED 159 663, 69p., PC-$4.82, MF-$.83)

No. 103: Fleisher, L. S., Jenkins, J. R., & Pany, D. *Effects on Poor Readers' Comprehension of Training in Rapid Decoding,* September 1978. (ERIC Document Reproduction Service No. ED 159 664, 39p., PC-$3.32, MF-$.83)

No. 104: Anderson, T. H. *Study Skills and Learning Strategies,* September 1978. (ERIC Document Reproduction Service No. ED 161 000, 41p., PC-$3.32, MF-$.83)

No. 105: Ortony, A. *Beyond Literal Similarity,* October 1978. (ERIC Document Reproduction Service No. ED 166 635, 58p., PC-$4.82, MF-$.83)

No. 106: Durkin, D. *What Classroom Observations Reveal about Reading Comprehension Instruction,* October 1978. (ERIC Document Reproduction Service No. ED 162 259, 94p., PC-$6.32, MF-$.83)

No. 107: Adams, M. J. *Models of Word Recognition,* October 1978. (ERIC Document Reproduction Service No. ED 163 431, 93p., PC-$6.32, MF-$.83)

No. 108: Reder, L. M. *Comprehension and Retention of Prose: A Literature Review,* November 1978. (ERIC Document Reproduction Service No. ED 165 114, 116p., PC-$7.82, MF-$.83)

No. 109: Wardrop, J. L., Anderson, T. H., Hively, W., Anderson, R. I., Hastings, C. N., & Muller, K. E. *A Framework for Analyzing Reading Test Characteristics,* December 1978. (ERIC Document Reproduction Service No. ED 165 117, 65p., PC-$4.82, MF-$.83)

No. 110: Tirre, W. C., Manelis, L., & Leicht, K. L. *The Effects of Imaginal and Verbal Strategies on Prose Comprehension in Adults,* December 1978. (ERIC Document Reproduction Service No. ED 165 116, 27p., PC-$3.32, MF-$.83)

No. 111: Spiro, R. J., & Tirre, W. C. *Individual Differences in Schema Utilization During Discourse Processing,* January 1979. (ERIC Document Reproduction Service No. ED 166 651, 29p., PC-$3.32, MF-$.83)

No. 112: Ortony, A. *Some Psycholinguistic Aspects of Metaphor,* January 1979. (ERIC Document Reproduction Service No. ED 165 115, 38p., PC-$3.32, MF-$.83)

No. 113: Antos, S. J. *Processing Facilitation in a Lexical Decision Task,* January 1979. (ERIC Document Reproduction Service No. ED 165 129, 84p., PC-$6.32, MF-$.83)

No. 114: Gentner D. *Semantic Integration at the Level of Verb Meaning,* February 1979. (ERIC Document Reproduction Service No. ED 165 130, 39p., PC-$3.32, MF-$.83)

No. 115: Gearhart, M., & Hall, W. S. *Internal State Words: Cultural and Situational Variation in Vocabulary Usage,* February 1979. (ERIC Document Reproduction Service No. ED 165 131, 66p., PC-$4.82, MF-$.83)

No. 116: Pearson, P. D., Hansen, J., & Gordon, C. *The Effect of Background Knowledge on Young Children's Comprehension of Explicit and Implicit Information,* March 1979. (ERIC Document Reproduction Service No. ED 169 521, 26p., PC-$3.32, MF-$.83)

No. 117: Barnitz, J. G. *Reading Comprehension of Pronoun-Referent Structures by Children in Grades Two, Four, and Six,* March 1979. (ERIC Document Reproduction Service No. ED 170 731, 51p., PC-$4.82, MF-$.83)

No. 118: Nicholson, T., Pearson, P. D., & Dykstra, R. *Effects of Embedded Anomalies and Oral Reading Errors on Children's Understanding of Stories,* March 1979. (ERIC Document Reproduction Service No. ED 169 524, 43p., PC-$3.32, MF-$.83)

No. 119: Anderson, R. C., Pichert, J. W., & Shirey, L. L. *Effects of the Reader's Schema at Different Points in Time,* April 1979. (ERIC Document Reproduction Service No. ED 169 523, 36p., PC-$3.32, MF-$.83)

No. 120: Canney, G., & Winograd, P. *Schemata for Reading and Reading Comprehension Performance,* April 1979. (ERIC Document Reproduction Service No. ED 169 520, 99p., PC-$6.32, MF-$.83)

No. 121: Hall, W. S., & Guthrie, L. F. *On the Dialect Question and Reading,* May 1979. (ERIC Document Reproduction Service No. ED 169 522, 32p., PC-$3.32, MF-$.83)

No. 122: McClure, E., Mason, J., & Barnitz, J. *Story Structure and Age Effects on Children's Ability to Sequence Stories,* May 1979. (ERIC Document Reproduction Service No. ED 170 732, 75p., PC-$4.82, MF-$.83)

No. 123: Kleiman, G. M., Winograd, P. N., & Humphrey, M. M. *Prosody and Children's Parsing of Sentences,* May 1979. (ERIC Document Reproduction Service No. ED 170 733, 28p., PC-$3.32, MF-$.83)

No. 124: Spiro, R. J. *Etiology of Reading Comprehension Style,* May 1979. (ERIC Document Reproduction Service No. ED 170 734, 21p., PC-$1.82, MF-$.83)

No. 125: Hall, W. S., & Tirre, W. C. *The Communicative Environment of Young Children: Social Class, Ethnic, and Situational Differences,* May 1979. (ERIC Document Reproduction Service No. ED 170 788, 30p., PC-$3.32, MF-$.83)

No. 126: Mason, J., & McCormick, C. *Testing the Development of Reading and Linguistic Awareness,* May 1979. (ERIC Document Reproduction Service No. ED 170 735, 50p., PC-$3.32, MF-$.83)

No. 127: Brown, A. L., & Campione, J. C. *Permissible Inferences from the Outcome of Training Studies in Cognitive Development Research,* May 1979. (ERIC Document Reproduction Service No. ED 170 736, 34p., PC-$3.32, MF-$.83)

No. 128: Brown, A. L., & French, L. A. *The Zone of Potential Development: Implications for Intelligence Testing in the Year 2000,* May 1979. (ERIC Document Reproduction Service No. ED 170 737, 46p., PC-$3.32, MF-$.83)

No. 129: Nezworski, T., Stein, N. L., & Trabasso, T. *Story Structure Versus Content Effects on Children's Recall and Evaluative Inferences,* June 1979. (ERIC Document Reproduction Service No. ED 172 187, 49p., PC-$3.32, MF-$.83)

No. 130: Bruce, B. *Analysis of Interacting Plans as a Guide to the Understanding of Story Structure,* June 1979. (ERIC Document Reproduction Service No. ED 174 951, 43p., PC-$3.32, MF-$.83)

No. 131: Pearson, P. D., Raphael, T., TePaske, N., & Hyser, C. *The Function of Metaphor in Children's Recall of Expository Passages,* July 1979. (ERIC Document Reproduction Service No. ED 174 950, 41p., PC-$3.32, MF-$.83)

No. 132: Green, G. M. *Organization, Goals, and Comprehensibility in Narratives: Newswriting, a Case Study,* July 1979. (ERIC Document Reproduction Service No. ED 174 949, 66p., PC-$4.82, MF-$.83)

No. 133: Kleiman, G. M. *The Scope of Facilitation of Word Recognition from Single Word and Sentence Frame Contexts,* July 1979. (ERIC Document Reproduction Service No. ED 174 947, 61p., PC-$4.82, MF-$.83)

No. 134: McConkie, G. W., Hogaboam, T. W., Wolverton, G. S., Zola, D., & Lucas, P. A. *Toward the Use of Eye Movements in the Study of Language Processing,* August 1979. (ERIC Document Reproduction Service No. ED 174 968, 48p., PC-$3.32, MF-$.83)

No. 135: Schwartz, R. M. *Levels of Processing: The Strategic Demands of Reading Comprehension,* August 1979. (ERIC Document Reproduction Service No. ED 177 471, 45p., PC-$3.32, MF-$.83)

No. 136: Anderson, R. C., & Freebody, P. *Vocabulary Knowledge,* August 1979. (ERIC Document Reproduction Service No. ED 177 480, 71p., PC-$4.82, MF-$.83)

No. 137: Royer, J. M., Hastings, C. N., & Hook, C. *A Sentence Verification Technique for Measuring Reading Comprehension,* August 1979. (ERIC Document Reproduction Service No. ED 176 234, 34p., PC-$3.32, MF-$.83)

No. 138: Spiro, R. J. *Prior Knowledge and Story Processing: Integration, Selection, and Variation,* August 1979. (ERIC Document Reproduction Service No. ED 176 235, 41p., PC-3.32, MF-$.83)

No. 139: Asher, S. R., & Wigfield, A. *Influence of Comparison Training on Children's Referential Communication,* August 1979. (ERIC Document Reproduction Service No. ED 177 493, 42p., PC-$3.32, MF-$.83)

No. 140: Alessi, S. M., Anderson, T. H., & Goetz, E. T. *An Investigation of Lookbacks During Studying,* September 1979. (ERIC Document Reproduction Service No. ED 177 494, 40p., PC-$3.32, MF-$.83)

No. 141: Cohen, P. R., & Perrault, C. R. *Elements of a Plan-Based Theory of Speech Acts,* September 1979. (ERIC Document Reproduction Service No. ED 177 497, 76p., PC-$6.32, MF-$.83)

No. 142: Grueneich, R., & Trabasso, T. *The Story as Social Environment: Children's Comprehension and Evaluation of Intentions and Consequences,* September 1979. (ERIC Document Reproduction Service No. ED 177 496, 56p., PC-$4.82, MF-$.83)

No. 143: Hermon, G. *On the Discourse Structure of Direct Quotation,* September 1979. (ERIC Document Reproduction Service No. ED 177 495, 46p., PC-$3.32, MF-$.83)

No. 144: Goetz, E. T., Anderson, R. C., & Schallert, D. L. *The Representation of Sentences in Memory,* September 1979. (ERIC Document Reproduction Service No. ED 177 527, 71p., PC-$4.82, MF-$.83)

No. 145: Baker, L. *Comprehension Monitoring: Identifying and Coping with Text Confusions,* September 1979. (ERIC Document Reproduction Service No. ED 177 525, 62p., PC-$4.82, MF-$.83)

No. 146: Hall, W. S., & Nagy, W. E. *Theoretical Issues in the Investigation of Words of Internal Report,* October 1979. (ERIC Document Reproduction Service No. ED 177 526, 108p., PC-$7.82, MF-$.83)

No. 147: Stein, N. L., & Goldman, S. *Children's Knowledge about Social Situations: From Causes to Consequences,* October 1979.

No. 148: Hall, W. S., & Guthrie, L. F. *Cultural and Situational Variation in Language Function and Use: Methods and Procedures for Research,* October 1979.

No. 149: Pichert, J. W. *Sensitivity to What is Important in Prose,* November 1979.

No. 150: Dunn, B. R., Mathews, S. R., II, & Bieger, G. *Individual Differences in the Recall of Lower-Level Textual Information,* December 1979.

No. 151: Gentner, D. *Verb Semantic Structures in Memory for Sentences: Evidence for Componential Representation,* December 1979.

No. 152: Tierney, R. J., & Mosenthal, J. *Discourse Comprehension and Production: Analyzing Text Structure and Cohesion,* January 1980.

No. 153: Winograd, P., & Johnston, P. *Comprehension Monitoring and the Error Detection Paradigm,* January 1980.

No. 154: Ortony, A. *Understanding Metaphors,* January 1980.

No. 155: Anderson, T. H., & Armbruster, B. B. *Studying,* January 1980.

No. 156: Brown, A. L., & Campione, J. C. *Inducing Flexible Thinking: The Problem of Access,* January 1980.

No. 157: Trabasso, T. *On the Making of Inferences During Reading and Their Assessment,* January 1980.

No. 158: McClure, E., & Steffensen, M. S. *A Study of the Use of Conjunctions across Grades and Ethnic Groups,* January 1980.

No. 159: Iran-Nejad, A. *The Schema: A Structural or a Functional Pattern,* February 1980.

No. 160: Armbruster, B. B., & Anderson, T. H. *The Effect of Mapping on the Free Recall of Expository Text,* February 1980.

No. 161: Hall, W. S., & Dore, J. *Lexical Sharing in Mother-Child Interaction,* March 1980.

No. 162: Davison, A., Kantor, R. N., Hannah, J., Hermon, G., Lutz, R., Salzillo, R. *Limitations of Readability Formulas in Guiding Adaptations of Texts,* March 1980.

No. 163: Linn, R. L., Levine, M. V., Hastings, C. N., & Wardrop, J. L. *An Investigation of Item Bias in a Test of Reading Comprehension,* March 1980.

No. 164: Seidenberg, M. S., Tanenhaus, M. K., & Leiman, J. M. *The Time Course of Lexical Ambiguity Resolution in Context,* March 1980.

No. 165: Brown, A. L. *Learning and Development: The Problems of Compatibility, Access, and Induction,* March 1980.

No. 166: Hansen, J., & Pearson, P. D. *The Effects of Inference Training and Practice on Young Children's Comprehension,* April 1980.

No. 167: Straker, D. Y. *Situational Variables in Language Use,* April 1980.

No. 168: Green, G. M., Kantor, R. N., Morgan, J. L., Stein, N. L., Hermon, G., Salzillo, R., Sellner, M. B., Bruce, B. C., Gentner, D., & Webber, B. L. *Problems and Techniques of Text Analysis,* April 1980.

No. 169: Green, G. M., Kantor, R. N., Morgan, J. L., Stein, N. L., Hermon, G., Salzillo, R., & Sellner, M. B. *Analysis of Babar Loses His Crown,* April 1980.

No. 170: Green, G. M., Kantor, R. N., Morgan, J. L., Stein, N. L., Hermon, G., Salzillo, R., & Sellner, M. B. *Analysis of "The Wonderful Desert,"* April 1980.