DOCUMENT RESUME

ED 236 575 CS 007 364

AUTHOR Olson, Gary M.
TITLE Composition and Comprehension of Simple Texts. Final Report.
INSTITUTION Michigan Univ., Ann Arbor.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE [83]
GRANT NIE-G-79-0133
NOTE 236p.; Several tables may be marginally legible.
PUB TYPE Reports - Research/Technical (143) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC10 Plus Postage.
DESCRIPTORS Cognitive Processes; Higher Education; *Reading Comprehension; Reading Processes; *Reading Research; Research Methodology; *Schemata (Cognition); Sentence Structure; *Writing (Composition); Writing Processes; *Writing Research
IDENTIFIERS *Reading Writing Relationship; Think Aloud Protocol

ABSTRACT

This report describes research that focused on the comprehension and composition of simple texts. The first section reviews the overall goals and theoretical perspectives of the project. The second section describes the following studies carried out during the project: analysis and extension of prior thinking-out-loud (TOL) data, TOL and reading time data for essays, controlled reading times, expectations and sentence integration, and writing processes. The third part lists the talks and papers that have resulted from the project. Appendixes include a scheme for coding think-aloud protocols, a coding scheme for descriptions of procedures, and copies of the following publications: (1) "Applying Knowledge of Writing Conventions to Prose Comprehension and Composition," (2) "Cognitive Aspects of Genre," (3) "Thinking Out-Loud as a Method for Studying Real-Time Comprehension Processes," (4) "Question-Asking as a Component of Text Comprehension," and (5) "The Role of Expectations in Sentence Integration." (HOD)

Final Report

NIE Grant G-79-0133

Composition and Comprehension of Simple Texts

Gary M. Olson
University of Michigan

Principal Investigator

TABLE OF CONTENTS

Final Report

NIE Grant G-79-0133

Composition and Comprehension of Simple Texts

Gary M. Olson
University of Michigan

Principal Investigator

This report describes the research carried out under NIE grant G-79-0133, during the period from September 15, 1979 to September 15, 1982. The first part reviews the overall goals and theoretical perspectives of the project. The second section describes a series of studies carried out during the project. The third part lists the talks and papers that have resulted from the project.

BACKGROUND

This program of research focussed on the comprehension and composition of simple texts. The tasks of composing text or understanding it are each exceedingly complex. The present research assumed that an adequate characterization of either activity requires that we consider the nature of the relation between the reader and the writer. Thus, we were interested in examining the reader's assumptions about how a text is written and how this affected comprehension. Similarly, we wanted to learn about how writers take account of their readers as they write or revise their text.

The primary features of the conceptualization that has motivated our research appears in Olson, Duffy and Mack (1980) and Olson, Mack and Duffy (1981). Rather than duplicate these discussions here, we will present a brief summary of the central points. The core idea was described in the preceding paragraph: the writer and reader communicate through a text according to a set of conventions about how a text ought to work. Such conventions govern all forms of social interaction, including all the different ways in which language is used. We have been particularly interested in those conventions of written communication that arise because of the special characteristics of writing when contrasted with speech. Table 1, from Olson et al. (1981), shows the major differences between speaking and writing.

On the basis of these differences, there are a number of conventions that regulate the process of written communication. Table 2, also from Olson et al. (1981), shows some of these. These principles apply to simple texts, such as the stories and essays that have been used in our research. By and large these principles are self-explanatory, but further discussion of them is presented in Olson et al. (1981).

Table 1
Characteristics of writing compared to speech

Characteristic	Description
1. Permanence	Writing persists through time, while speech is highly transient.
2. Detachment	Both the content and form of written language is divorced from the immediate context in space and time.
3. Absence of feedback	Writing is a one-way process, without feedback.
4. Nonspecificity	Writing is typically addressed to a general audience rather than a specific individual.
5. Tellability	The topic of the text deserves to have the trouble taken to write it up.
6. Organization	Written language is much more planned and organized than speech.
7. Formality	The language of writing tends to be more formal than speech.
8. Economy	Written language has less redundancy than spoken language.
9. Greater precision and detail	Written language can develop a topic in greater detail
10. Greater complexity and abstractness of subject matter	Written language can develop more complex ideas.

These principles guide the activities of both the writer and the reader. In most of the work carried out under this grant, we have studied the strategies of readers, using a thinking-out-loud methodology (see Olson, Duffy and Mack, 1984, for a detailed discussion) to reveal exactly how readers go about extracting the meaning of a text, using these principles as guidance. In a study carried out during the last year of the grant we examined the activities of writers, studying how texts written to convey a particular content were assembled and later revised. The findings of these investigations will be summarized in the following sections.

RESEARCH

The collection of empirical data on comprehension and composition was the central focus of activity during this grant. Because of the scope and complexity of the data collected, not all analyses and write-ups have been completed. In each part of this section, a general overview will be given of the type of data that was collected and the rationale for the work. Where detailed analyses have been completed, a summary will be presented and publications or talks that have resulted will be listed. Each subsection will conclude with a discussion of the plans for further analysis and publication of the data we have on hand.

As planned, the bulk of the research completed during this grant was on comprehension. However, during the third year of the project we began work on composition that grew out of our previous work on

Table 2
Conventions of composition for simple texts.

Convention	Description
1. Purpose	The writer has one or more specific purposes in mind.
2. Balance of novel and familiar elements	The text contains a balance of familiar and novel elements.
A. Familiarity	The text makes contact with things the reader knows about.
B. Novelty	The text contains new or distinctive elements.
3. Underlying organization	The text is based on an underlying structure that is appropriate and well-formed for the particular genre.
A. Focus	There is one main line of development.
B. Overall plan	There is an overall plan that provides a well-formed organization for the propositions of the text.
(i) Coherence	Each individual proposition fits in a well-ordered way into the general plan.
(ii) Completeness	The overall organization has a closed or optimal structure.
C. Conventional world	The text is based on a world that is similar but not identical to the real world.
4. Surface organization	The writer has a surface plan for presenting the surface propositions of the text.
A. Omniscience	At any point in the text the writer knows where it is going.
B. Audience	The text is written with an audience in mind.
C. Scaffolding	The writer presents enough supporting material in the text so that the reader has sufficient background to be able to reconstruct propositions where necessary.
D. Segmentation	The text is organized into surface chunks.
E. Connectivity	The writer provides sufficient textual information so that the reader can make all the necessary inferential connections among the elements of the text.
F. Economy	Everything that is presented has a purpose within the surface plan.
G. Orderly flow	The sequence of segments and propositions in the surface of the text is principled.
H. Language	There are conventions for the surface language of each genre.
(i) Signalling	Appropriate surface signals will be used to mark transitions, etc.
(ii) Level	Appropriate levels of language will be used
I. Genre-specific conventions	In addition to the general conventions listed above particular genres will often have specific text conventions governing the surface form.

comprehension. The specific research carried out is described in the next two sections.

STUDIES OF COMPREHENSION

1. ANALYSIS AND EXTENSION OF PRIOR TOL DATA

Prior to submitting the proposal which led to funding of this project we had collected thinking-out-loud data for four stories. The

focus of this work was the question: What is a reader doing while reading a simple story? To get ideas, we had a number of subjects think out loud as they read simple stories. Since many details of the methodology have been discussed in Olson, Duffy and Mack (1984) and many of the major findings have been presented in Olson, Mack and Duffy (1981), only the highlights of the results of this work will be presented here prior to describing what additional work we have done.

The thinking-out-loud data provided two kinds of information about the processing of simple stories. Analysis of the general content of the protocols allowed us to construct a general picture of how an intelligent reader approaches the task of reading a story. As described in Olson, Duffy and Mack (1980) and the two other papers just cited, readers of stories engage in large amounts of predictive, problem-solving behavior as they read through a story. They possess a large amount of general world knowledge and specific story knowledge that they apply in a predictive fashion as they read through the text.

A central question posed in our proposal is the extent to which this portrait of story understanding actually holds for readers who are not thinking out loud and for texts other than stories. A variety of approaches were used to address this question. One was to look at the relationship between thinking-out-loud data and other measures of processing, such as reading time. Multiple regression analyses of reading time data, reported in Olson, Mack and Duffy (1981), showed that for well-formed stories there was a relationship, with silent readers reading more slowly at the same places where thinking-out-loud readers talked more (with the obvious confounding effect of sentence length controlled). This finding has been central to much of our subsequent work, and has justified for us the value of thinking-out-loud data for the study of comprehension (see a detailed discussion of the pros and cons of this method in Olson, Duffy and Mack, 1984). A second approach was to collect a comparable set of data for a genre other than simple stories. This work is described in section 2. A third approach was to collect new protocol data for stories and essays, using more focussed, analytic tasks. This is described in section 3. Finally, several experimental studies of specific predictions that emerged from this view of comprehension were pursued, the best example being Duffy's dissertation research, described in section 5.

We recognized that the original set of thinking-out-loud data we had collected were exceedingly rich, and so during the course of this project we have devoted some of our energies toward their further analysis. In particular, we wanted to get a deeper, richer picture of exactly what these readers were doing. To this end, we developed a scheme for coding the content of these protocols, and have completely coded the protocols for one of our four stories, Lentil.

The coding scheme is based on a simple principle. Each utterance in a thinking-out-loud protocol says something about some part of the text. Thus, two features are coded for each protocol segment: that part of the text that is referred to, and what was said about it. The first of these we refer to as an Attention code, the second as an Operation code. The coding scheme is shown in detail in Appendix A.

Considerable time was devoted to developing and refining this scheme, and establishing its reliability. We required that two trained scorers agree on their codings at least 80% of the time. For purposes of final coding, the disagreements were resolved by discussion between the coders. The original thinking-out-loud data collected for 12 subjects for the story Lentil (see Olson et al., 1981) have been completely coded and checked, and a series of analyses of these data are planned but not yet completed.

This work is still incomplete. It is time-consuming work, and until we are convinced it is sufficiently useful, we do not want to code the rest of our protocols. We plan to complete our analysis of the Lentil data and then make a decision about the treatment of the remainder of this rich data base.

2. TOL AND READING TIME DATA FOR ESSAYS

An early concern of ours was the representativeness of our findings with stories. Would a similar picture of comprehension emerge for other text genres? Early in this project we collected thinking-out-loud, reading time, and recall data for academic essays. We selected four essays, and for each one we had two versions. The portrait of processing we found and the detailed quantitative analyses of these data were quite different than what we found with stories. As summarized in Olson, Mack and Duffy (1981), while the reader of a story approaches the text in a predictive, prospective fashion, focussing on what is coming up, the readers of our essays approached their texts in a retrospective fashion, fitting in what they were currently reading with what had come before but making only the most general, vague predictions about what they were going to be reading (and making these largely because we asked them to). Further, unlike the stories, we found no relationship between the TOL data and silent reading times. The details of much of this work appeared in Olson, Mack and Duffy (1981).

There are additional details of these data that still need to be examined. Duffy, Olson, Mack, Vincent and Eaton (1982) reported some analyses of reader's strategies while reading essays, and these analyses will be expanded and then reported in a new paper. We also want to explore the usefulness of a scheme like that shown in Appendix A applied to essays.

3. NEW DATA FOR STORIES AND ESSAYS

The general TOL data we collected for stories and essays (just reviewed in sections 1 and 2) revealed to us a number of further questions that we decided to pursue with more focussed TOL tasks (see a discussion of the rationale for the different types of TOL tasks in Olson, Duffy & Mack, 1984). One focus was predictions. Since our story TOL data had shown a clear indication of the central role of predictions in story processing, and we had found such a marked contrast between stories and essays in this, we collected new TOL data in which we had subjects only give predictions after reading each sentence. Another

interest stimulated by our previous TOL data was on questions asked by subjects as they were reading. This in turn led us to collect data in which subjects asked questions following each sentence of a story or an essay. Of the four types of new data collected (story/essay crossed with prediction/question), only the question-asking data for stories have been analyzed sufficiently to present the findings.

Question-Asking for Simple Stories. In its primary mode of use, a question is a device for seeking new information that is to be related to an existing knowledge structure. When to ask a question, and exactly what to ask, are both symptomatic of the status of the knowledge structure at issue, as well as, no doubt, the general intelligence of the asker. We have all encountered the person (often ourselves!) who indicated they did not know enough about a topic to ask a question about it. Thus, intuitively, there is a link between one's knowledge or understanding of a topic and the ability to ask a question about it (e.g., see Miyake & Norman, 1979).

There is another connection between questions and comprehension. Educators and researchers have long suspected that approaching the comprehension of text with either general or specific questions in mind might facilitate understanding. There is a sizable research literature on this role of questions in understanding text (e.g., Anderson & Biddle, 1975; Frase, 1975). Questions of this type focus the reader's attention on exactly those pieces of information that are important to understanding what the text is about. Since one of the problems faced by the reader is selecting the most relevant or important information from a text, appropriate questions can serve as a guide for this important process.

These two uses of questions in relation to understanding have an important relationship. Questions asked about a text are both an indication of having understood what has been read and a guide to the further understanding of what is about to be read. This suggested to us that questions asked by a reader while reading a text might be an especially informative kind of data for monitoring the reader's understanding of the text.

In our earlier work (Olson, Duffy & Mack, 1980, 1984; Olson, Mack & Duffy, 1981), one of the things we noticed subjects doing while thinking out loud during reading was asking questions. The kinds of questions people asked and the places they asked them seemed to us indicative of important comprehension processes. This led us to conduct a specific study on the relationship between on-line question asking and comprehension. We shall report a few highlights of this study here. A more complete report of it will appear in Olson, Duffy, Eaton, Vincent and Mack (in preparation).

Let us summarize the general rationale for this study. The kinds of considerations we have sketched led us to believe that questions asked by subjects during the reading of a simple text would be diagnostic of important comprehension processes. It seemed plausible to assume that each sentence encountered in a text raises certain questions in a reader's mind and answers other questions raised by earlier

sentences. We wanted to explore this supposition in more detail by collecting data on the kinds of questions readers ask following each sentence in simple stories.

This study used four tasks. The primary task was one in which readers asked questions after reading each sentence in the story. In another task a different group of subjects read the same stories silently while we timed their reading. These same subjects later recalled the stories. Finally, another group of subjects rated the importance of the constituents of the story. Four short simple stories (maximum length was 41 sentences) were used as texts. They were all children's stories or simple folktales, and all were well-formed.

To better understand the results, a somewhat more detailed description of the four tasks is necessary:

1. Question-asking. All four stories were presented to 9 subjects. Each sentence in the story was typed on a card, and the subject worked his or her way through the deck of cards, asking questions that were raised in his or her mind as a result of having read that particular sentence. The subject was told to imagine that the story's author was present, and that the author was willing to answer any questions the reader had about the story at that point, except for the obvious question of what happens next. The subject was allowed to spend as much time on any sentence as he or she desired, but was asked not to reread any previous sentences or to look ahead. The questions were tape recorded and later transcribed. The number of questions asked for each sentence was tallied and pooled over subjects. In addition, the questions were classified in various ways.

2. Reading times. Sentence-by-sentence reading times were collected from 20 subjects. At the end of each story subjects wrote a brief (3 to 5 sentences) summary of the story.

3. Recall. The same 20 subjects were asked to recall the stories they had just read. They were presented with a brief descriptive title for each story, and were given unlimited time to try to recall as much as they could. They were asked to recall exact words, but were encouraged to guess if they could not remember exact words. Recall was scored by first doing a propositional analysis of each story and then matching the subject's recall against this, using a gist criterion.

4. Importance. Seventeen subjects read each story and crossed out the 50% of the words, phrases, or sentences in the story they felt was least important. For each sentence in each story the proportion of words left in averaged over subjects provided a measure of the relative importance of that sentence.

It is useful to have a better picture of what the question-asking data look like. Table 3 shows typical questions for the first sentence of one of the stories. These questions are grouped into those asked by two or more subjects and those that are idiosyncratic to one subject. Of course, we were also interested in the sentence-by-sentence variation in the questions asked. Figure 1 shows the total number of questions

asked for each sentence in each of the four stories. With the possible exception of EMERALD, there is noteworthy variation in the number of questions asked from sentence to sentence. In EMERALD, there were a large number of questions at the beginning and then a fairly flat distribution of questions thereafter. Keep this difference in mind, because EMERALD will not follow the pattern of other stories in some of our later analyses.

The first issue we addressed was whether the question-asking task is related to the reading times. We examined this by looking at the relationship between the total number of questions asked for each sentence in a story and the average reading time for each sentence for those subjects who were reading silently. The expectation was that sentences which elicited a lot of questions would be especially salient during real-time processing, and therefore would be read more slowly by subjects who were reading silently. This hypothesis was confirmed. We conducted multiple regressions in which the average reading time per sentence was the dependent variable, and the predictor variables were sentence length, total number of questions, serial position, and importance. Only sentence length and number of questions emerged as significant predictors of reading time. In this analysis all four stories were entered, with story as a variable. There are two types of questions that occur: those that are asked by several subjects, and those that are idiosyncratic. We next asked whether these two types of questions contributed differentially to this outcome. The answer was no. A multiple regression with number of questions asked by two or more persons and idiosyncratic questions entered separately showed that both emerged as significant predictors. Table 4 shows the details of these analyses.

When we carried out multiple regression analyses for each story individually, the results mirrored the overall analysis. In these regressions we included as predictors idiosyncratic questions and questions asked by two or more persons as well as total number of questions asked. For three of the four stories, at least one of these question counts emerged as a significant predictor of reading time (in addition to number of syllables). The exception was EMERALD, for which the question data provided no significant predictor. As mentioned earlier, EMERALD was the story that showed little variation in number of questions asked across sentences.

So, number of questions asked accounts for a significant portion of the variance in sentence-by-sentence reading times. We next asked what relationship the question-asking task has with recall. And the answer was very simple: none. Table 5 shows the outcome of a multiple regression carried out on recall scores, and reveals that rated importance and serial position emerged as significant predictors of recall, while number of questions asked did not. This pattern is similar to other data which indicate that importance predicts recall (Meyer, 1975; Kintsch, 1974). Importance is not necessarily immediately perceived, but may result from having most or all of the final memory representation of the text. We conclude from this that the information being revealed by the question-asking task is more closely associated with the activities that occur during comprehension than with the form

Table 3

Sample Questions from The Selling of the Cow

Sentence 1:

"Once there was a man named Cromer who lived on a farm that was way up on the side of a hill."

Questions asked by 2 or more subjects:	# Subjects
1) Who is Cromer?	2
2) What is Cromer like?	3
3) Did Cromer live alone?	5
4) When did this story take place?	5
5) Where was the farm?	4
6) Where was the hill?	4
7) Why was the farm on a hill?	2
8) How far up the hill was the farm?	2
9) How high was the hill?	3
10) What kind of farm was it?	4
11) What will happen to Cromer?	2

Idiosyncratic questions asked by only 1 subject:

1) Does the fact that he lives on a farm have any significance?
2) Does he farm for a living?
3) Does he have another vocation?
4) Is Cromer married?
5) How old is Cromer?
6) How far away were Cromer's nearest neighbors?

7) Why did Cromer like to live on a farm?

8) Are they going to roll something down the hill?

9) Did a lot of the dirt wash off the side of the hill so that Cromer couldn't have his crops?

10) What was Cromer's first name?

11) Was that Cromer's first name?

12) Then what was Cromer's last name?

13) Did Cromer have more than one name?

14) What kind of name is Cromer?

15) What does Cromer mean?

16) What nationality is Cromer?

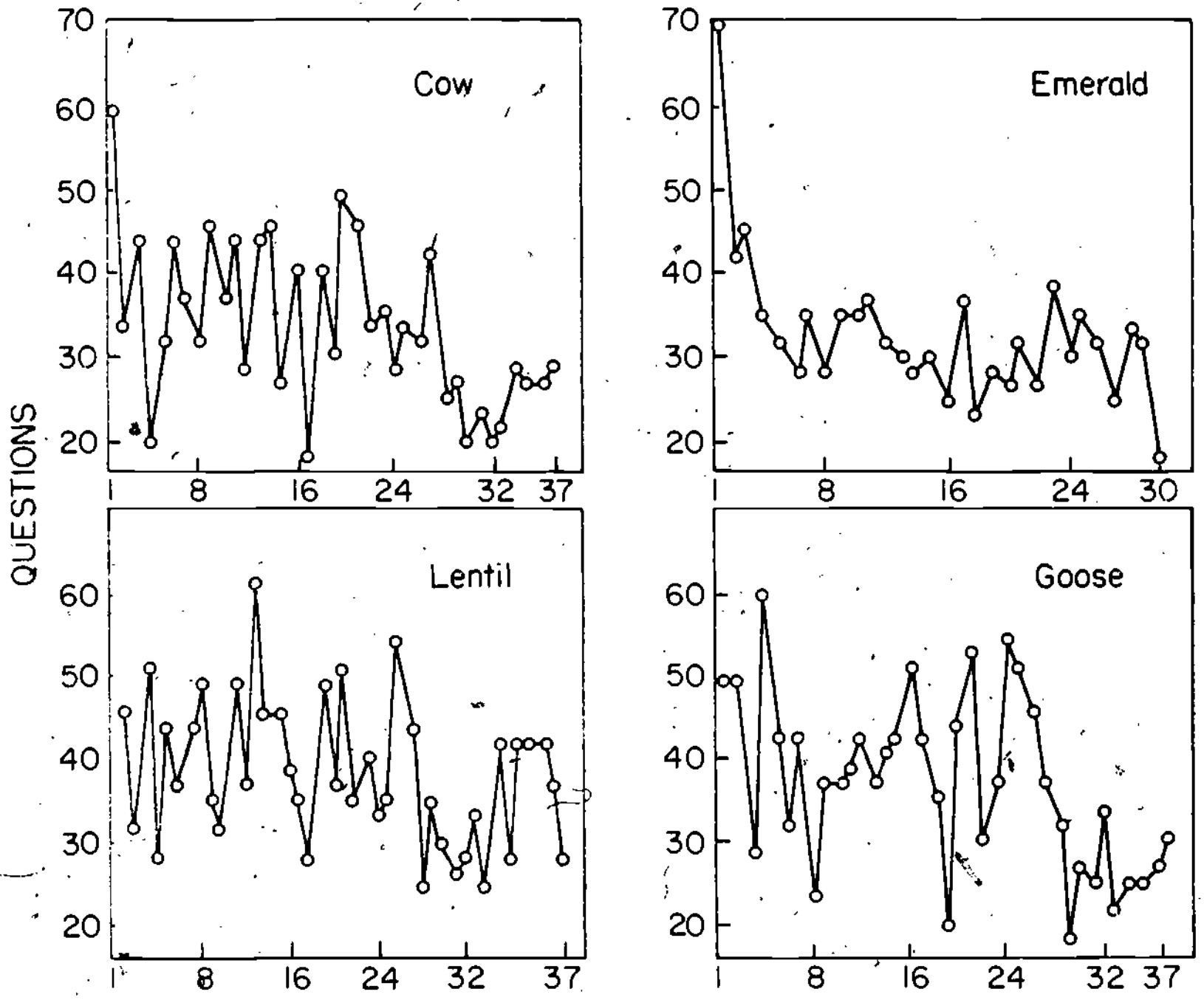

Cow
Emerald
Lentil
Goose
QUESTIONS
SENTENCE NUMBER
70
60
50
40
30
20
1
8
16
24
32
37
30

Table 4

Multiple Regression Analyses of Reading Time in Question-Asking Experiment

	Regression Coefficient	Significance Level	Cumulative R^2
1. Predictors selected:			
Sentence length	130.22	.0001	.589
Total number of questions asked	26.44	.0001	.640
Predictors not selected:			
Serial position of sentence			
Importance			
2. Predictors selected:			
Sentence length	130.81	.0001	.589
Idiosyncratic questions	40.27	.0004	.626
Number of questions asked by two or more subjects	87.42	.0015	.652
Predictors not selected:			
Serial position of sentence			
Importance			

Note: Forward stepwise regression. dependent variable=mean reading time per sentence.

of the final memory representation constructed as a result of comprehension.

This basic result confirms our initial supposition that the question-asking task would tap an aspect of what is going on in the skilled reader's mind while reading. The obvious question, of course, is what is it tapping? It is unlikely that a reader who is reading silently is actually asking questions while reading. Rather, we believe that the question-asking task taps the kinds of informational needs a reader encounters while proceeding through a text. As each sentence is understood and added to a growing representation of the story, the reader revises and elaborates the set of information still needed to have the developing story make sense. These informational needs interact with what is presented in the next sentence to generate a new set of informational needs--or, if you will, a new set of questions--that guide the reader's comprehension through the succeeding parts of the text.

We have conducted a number of other analyses of these data that will be discussed in Olson, Duffy, Eaton, Vincent, and Mack (in preparation). We have categorized the questions to see if certain types are more important than others. So far, the categories we have examined have not shown any differences. We have also looked to see whether or not questions asked are later answered by the story, and there are interesting relationships here. Many questions <u>are</u> in fact answered, though it varies somewhat by type. However, the number of questions <u>answered</u> by a particular sentence does <u>not</u> predict reading time or recall. We have looked at the information tapped in the question, and find that questions which are derived from new information contained in the current sentence are especially important in predicting reading times. These and other details of these data are interesting and important, and will be reported on fully in Olson et al. (in preparation).

The main findings of this study strongly suggest that the question-asking task is a useful indicator of processes which may be an important part of comprehension. The number of questions asked by subjects as they read through a story correlates with the amount of time spent on that sentence by other readers reading silently. Keep in mind that this result is with the obvious effect of sentence length removed. But number of questions does not correlate with recall. Thus, question-asking seems more closely related to the real-time processes that occur during reading than to the final product of comprehension that remains when reading is completed.

How general are these findings? We do not yet know. We have question-asking data for academic essays, but have not yet analyzed them. This will be done in the coming year. Further, we will also analyze the prediction data we have collected for both stories and essays.

Table 5

Multiple Regression Analyses of Recall in Question-Asking Experiment

	Regression Coefficient	Significance Level	Cumulative R^2
Predictors selected:			
Importance	341.12	.0001	.235
Serial position of sentence	-2.94	.0024	.283
Predictors not selected:			
Sentence length			
Total number of questions asked			

Note: Forward stepwise regressions, dependent variable=proportion propositions recalled per sentence.

4. CONTROLLED READING TIMES

In our research we have assumed that the distribution of readings times across sentences is related to optimal reading strategies. Specifically, we have assumed that more time is devoted to those parts of texts that require or allow inferential and integrative processing. This assumption was tested by presenting the sentences of a text for experimentally determined exposure durations. Free reading data were collected for five short stories (range of lengths: 30 to 41 sentences). These free reading times were regressed against sentence length, and two experimental conditions were created using the residuals of this regression. In the Uniform condition, the exposure times for individual sentences were determined purely on the basis of sentence length, using the values calculated in the regression analysis. In the Congruent condition, the average reading times for individual sentences in the free reading condition were used. In both experimental conditions the total study time for each story were equal. What varied was how the time was allocated to individual sentences.

Each subject read all five stories and wrote a short summary of the story after reading each one. The first story was a practice one, and the remaining four were the materials of primary interest. One group of 21 subjects provided the free reading data, and a second group of 8 subjects provided controlled reading times (either Uniform or Congruent). For the latter subjects, half their stories were in each of the two experimental conditions. The practice story was always presented in the congruent condition. After reading all five stories, subjects were asked to recall the four primary stories in as much detail and with as much exact wording as possible. The order of recall was identical to the order of presentation. The story to be recalled was cued by a short title. The text of each story was given a propositional analysis as outlined in Turner & Greene (1977), and a gist criterion was used in scoring the recall.

Subjects who studied the texts in the free reading condition recalled 26.5% of the propositions. In the controlled reading conditions, 33.9% were recalled in the Congruent case and 32.9% in the Uniform case. The later difference was not significant.

The absence of an effect for Congruent vs. Uniform presentation condition led us to design a new study. Several factors occurred to us as relevant. First, perhaps processing of individual sentences is more flexible for some text types than for others. Second, perhaps some subjects are less disrupted by a non-optimal distribution of time than others. Therefore, we designed a new experiment in which subjects of varying known degrees of reading skill were given simple essays to read. A large pool of subjects were given the Nelson-Denny reading test, and from this pool subjects were assigned to high and low reading ability groups based on their scores. Subjects in the free reading group provided profiles of reading times that were used to construct the reading times used in the Uniform and Congruent conditions. Free recall of the text was once again the dependent variable.

At the present time, the second study is still being run. Pending its outcome, further studies along this same line may be conducted.

5. EXPECTATIONS AND SENTENCE INTEGRATION

This series of studies constituted Duffy's dissertation (Duffy, 1983), and has recently been submitted for publication. One endeavor within research on reading comprehension is to characterize how sentences are integrated into a coherent structure as the reader progresses through a text. Current approaches to sentence integration focus on backward search and inference processes to model the integration process. This research examined the role of predictive processing in sentence integration. The research had two goals: (1) to test the hypothesis that skilled readers regularly make predictions as they read, and (2) to constrain a model of predictive processing by providing information about the processing consequences of having formed the prediction. The term "prediction" is used to refer to several types of forward inferences, from the minimal prediction of the general topic of the next sentences to a specific content prediction of what will happen next.

Three experiments were carried out. In all three, stimuli were narrative text fragments which varied in the degree to which they generated a strong prediction at the end of the fragment. In the first experiment subjects were faster to respond that a target sentence was related to the text when it followed a High Expectation (HiE) text than a Low Expectation (LoE) text. Subjects were also faster to respond that a target sentence was unrelated when it followed a HiE text than when it followed a LoE text. In the second experiment, when subjects read a target sentence which conveyed the next event in a script, they took longer when they had an incorrect expectation than when they had no expectation. The third experiment failed to provide evidence that readers were forming a <u>specific</u> content prediction as they read.

The results show that readers generate predictions as they read. Furthermore, these predictions are generated selectively (not for every sentence), and they have processing consequences. A correct prediction can facilitate comprehension of the sentence where the prediction is fulfilled. An incorrect prediction can interfere with the processing of a sentence which violates the prediction. This pattern suggests that predictions are allocated some attention when generated and become involved in the processing of subsequent sentences.

These experiments fit nicely with our earlier work on thinking out loud during reading. That earlier work showed that readers readily made predictions while reading simple stories, and that the frequency of such predictions correlated with the silent reading times of different subjects. The Duffy dissertation results show even more clearly that readers of simple stories make predictions that have consequences for sentence-by-sentence processing.

A STUDY OF COMPOSITION

As discussed in the introduction, a significant aspect of good writing is the ability to take into account the reader's perspective. The writer's task is to communicate something to the reader in such a way that the reader can learn from the text in an orderly and efficient way. The absence of immediate feedback is a significant handicap for the writer. In conversation, the listener usually signals when the speaker has gone astray or is unclear. But the writer must put together a complete text, a complete set of thoughts, without such feedback.

On the basis of our prior work on real-time processes in comprehension (Olson, Duffy, & Mack, 1980, 1984; Olson, Mack & Duffy, 1981), we were interested in exploring the usefulness of process feedback to writers. We had found, both in our formal research and in informal observations, that the information given by a reader thinking out loud while reading contained much information that appeared to be useful to the writer. The think-aloud data gave precise feedback about what readers were doing, what features of the text they were reacting to, and what assumptions about the text and the writer they were making. It occurred to us that this information might be useful to a writer. In order to make the feedback as clear as possible to the writer's we chose material for communication that would have visible correlates of comprehension, namely, simple procedures. By having readers both perform the procedure and think out loud, the writers ought to have the most information possible about how well the readers were understanding the text.

Another motivation for this work was to follow up on Miller's (1980) important study of the natural language description of procedures. In his study, college students were asked to write descriptions of how to do a file manipulation problem (looking up or modifying information in the personnel files of a hypothetical company). The type of problem was varied across groups of subjects. The texts were examined from a number of perspectives in order to learn how computer-naive people would use natural language to describe a procedure similar to those that are typically programmed. Miller (1980) reported a number of details of his data, but only some of his broad conclusions will be reviewed here. Perhaps the most important finding was that people relied heavily on the fact that those who would be reading their descriptions were knowledgeable and intelligent. Much that was relevant to communicating the procedure was left implicit in their texts. For example, conditional statements were typically incomplete. The reader was told what to do when the condition was satisfied but nothing about what to do when it wasn't. In general, statements having to do with the control of action were implicit or missing. Similarly, references to other portions of the text were made implicitly rather than with explicit labels or directions. These are striking departures from procedure descriptions for a computer, where everything must be made very explicit.

These concerns converged to yield the following study. College students were taught two procedures. They then wrote a description of each procedure that could be used to teach another student how to do it.

Half of the writers had a series of readers read their texts, both thinking out loud and actually trying to perform the procedure. These reader sessions were videotaped, and later each writer watched three different readers talk about his or her texts while reading them. With the aid of this feedback, the writers revised their texts. A comparison group of writers revised their texts after the passage of an equivalent amount of time but without any feedback. A series of global ratings of the texts were then obtained from a set of judges who had been taught the procedures. Several general considerations motivated the analyses: Is process feedback of this sort useful for revising such texts? What are the properties of effective natural language descriptions of procedures, and how do they differ from ineffective ones?

Since this research is only just now being written up for publication, a more detailed description now follows.

Subjects. There were three different sets of subjects. The Writers consisted of 24 college students drawn from a standard subject pool and paid for their participation. There were 12 Experimental and 12 Control Writers who were roughly matched on sex, age, class in school, major, and computer experience. Data were also collected from 3 extra Experimental Writers, to be used as described later. The Readers consisted of 45 (36 for the 12 Experimental writers and 9 for the 3 extra ones) college students drawn from the same subject pool as the Writers. Finally, the Raters were 12 graduate students and postdocs recruited primarily from the Human Performance Center. None of the subjects knew the purpose or design of the study.

Tasks. In designing this study we considered a wide range of potential procedures. Some were procedures that some subjects would be expected to know and others would not (e.g., knitting), while others were ones that few would know. In the end, we chose the latter type of procedure since we did not want to preselect subjects on whether or not they knew the procedures. Thus, the Writers and Raters had to be taught these procedures and, of course, the Readers learned them from the texts.

The two procedures were called Card Sort and Fix. Card Sort was derived from a standard sorting algorithm in computer science (Knuth, 1973). It consisted of a set of steps through which an array of cards with numbers on them could be sorted into increasing numerical order. Fix was a dice game we invented. A die is thrown five times and a total score is computed on the basis of an algorithm that has certain analogies to the scoring of bowling.

Training tapes. A training video tape was made for each of the two procedural tasks. Since the Writers and Raters were to learn these tasks on the basis of these tapes, we wanted verbal commentary to be minimal so as not to provide them with a linguistic basis for their task. Thus, each procedural task was taught with as little language and as much gesturing and demonstration as was feasible. The same research assistant performed the task in both the tapes. Each tape was approximately 8 minutes long.

Design and procedure. The basic design of the study was a 2x2x2 factorial, consisting of Writer's Condition (Experimental vs. Control), Version of Text (Original vs. Revised) and Procedural Task (Card Sort vs Fix).

Each Writer participated in an individual training and writing session with each of the two procedural tasks. In each of these sessions the Writer first watched the training tape as many times as he or she wanted. The Writer then worked a series of problems until the experimenter was satisfied that the procedure was fully understood. Next the Writer wrote instructions for the task that could be understood by a naive college student. They were allowed as much time as they wanted. Typically, it took from 60 to 90 minutes to complete the training and the writing of the original version of the procedural description. The Card Sort and Fix tasks were done in separate sessions a few days apart, counterbalanced for order.

One to two weeks later, each Writer returned for a second session. The Control Writers were given a typed copy of their original texts, and were asked to reread their description and revise it. No specific advice was given about how to revise. Each Experimental Writer was shown three video tapes of Readers thinking out loud while reading that Writer's text. These Writers were told to use the information contained in these tapes to help them revise their text.

Each of the 45 Readers was run individually. Each Reader read a Card Sort and a Fix text, but from two different Writers. For each task, the Reader was given the materials needed for that task, and the text, and was told to try to learn how to do the task from the instructions given. The Reader was told to read the text out loud and to keep their finger pointed at the portion of the text they were reading. They were also asked to think out loud about their understanding of the text, to report what they were thinking about and doing and to make comments about the text or about their understanding of it. They were told that a video tape of their session would be shown to the Writer as feedback about how effective the text was, so that they should make comments they felt would be helpful to the Writer.

Each text was given a global rating by Raters who had been taught the procedural tasks. Each Rater received 12 pairs of texts, each pair consisting of the Original and Revised version for a particular Writer. Unknown to the Rater, half of the pairs were Experimental Writers and half Control. A pseudo-Latin square procedure was used to assign a different set of 12 Writers to each Rater such that each Writer was evaluated equally often across all Raters (each Writer had six ratings per procedural task). The texts of the three extra Writers were given as the first three pairs for all Raters to minimize contamination of the ratings by start-up effects. Thus, each Rater rated a set of 15 pairs of texts for each procedural task. The ratings for the two tasks were collected approximately a month apart.

For each pair of texts the Rater was asked to select which one most effectively communicated the procedural task. Since they had been trained on each task, they made this judgment from the perspective of

someone who already knew the procedure. Once they had selected the most effective member of the pair, they rated on a 1-7 scale (1=hard, 7=easy) how easy it was to make the decision. For the Fix task, which was evaluated about a month later than the Card Sort task, two additional judgments were made. One was a rating of how different the two texts were (1=very similar, 7=very different). The other was an absolute judgment on a 1-7 scale (7=very effective) of how effective each member of the pair was on its own. These judgments were added after some of the preliminary data for the Card Sort task had been evaluated.

Results. The analyses to be reported here fall into two broad categories. The first set address the question of whether the feedback to the Experimental Writers produced better revisions than those done by Control Writers. The second set compares the properties of the best and the worst of the descriptions, using a categorization scheme we have developed.

To give a rough idea of the size of these texts, Table 6 shows the average number of words for the texts in each of the eight cells of the design. Note that on balance the revisions were neither longer nor shorter than the originals. This is because some writers revised in ways that produced longer texts, whereas others did so in ways that produced shorter ones. Further, these average changes did not differ by experimental condition or procedural task. However, Table 7 sheds further light on these effects. This Table shows the mean differences between Original and Revised texts, both algebraically and absolutely, as well as the standard deviations of these differences. A clear picture emerges. Though on average both Experimental and Control Writers produced no net changes in length, because some had longer revisions and some shorter, the magnitude of these changes was larger for Experimental Writers. This is shown by the significantly larger absolute changes and by the significantly larger standard deviations for both measures for the Experimental Writers.

Another way to look at change would be to examine what proportion of the original text was changed on revision and --a somewhat different measure--what proportion of the revision appeared in the original. However, at this point these more difficult to compute measures have not been obtained.

On balance, then, the Experimental Writers made more changes than the Control Writers. Did they produce more effective texts in doing so? This can be examined in several ways. Table 8 shows the proportion of Raters who chose the revision as the more effective text for each condition and procedural task. These were evaluated by computing the proportion of Raters who chose the revision as the better of each Writer's text. Though there is a trend in the direction of the Experimental revisions being consistently more effective, this trend was not statistically significant. Similarly, the difference between the two procedural tasks was not significant.

Were the revisions on average better than the original texts? This is evaluated by the extent to which the revisions were selected more often than the chance value of 50%. The two cells for the Fix task did

Table 6

Mean Number of Words per Text

Task	Control		Experimental	
	Original	Revised	Original	Revised
Fix	333	351	366	386
Card Sort	575	570	444	425

Table 7

Mean Difference in Length Between Original and Revised Texts

Task	Signed Difference		Absolute Difference	
	Control	Experimental	Control	Experimental
Means				
Fix	18.3	19.9	18.7	40.1
Card Sort	-4.3	-19.5	21.9	46.5
Standard Deviations				
Fix	24.8	47.8	24.5	31.1
Card Sort	34.3	79.9	26.0	66.7

Note: Entries are mean words revised minus mean words original

not differ significantly from chance ($.05 < p < .10$) but both cells for the Card Sort task did ($p < .05$). Thus, the revisions were generally improved over the originals, but not differentially as a function of experimental condition--at least within the statistical power of this preliminary study.

Another way to look at this is to examine the overall ratings of text effectiveness that were obtained only for the Fix task. These are shown in Table 9. Statistical evaluation revealed that neither condition nor version affected these ratings.

Table 8

Proportion of Raters Choosing Revision as Better

Task	Condition	
	Control	Experimental
Fix	.62	.65
Card Sort	.70	.81

Table 9

Rated Text Effectiveness for Fix Task

Condition	Version of Text	
	Original	Revised
Control	4.3	4.5
Experimental	4.1	4.1

Note: Rating scale was 1-7, with 7 being most effective.

So, to summarize these preliminary analyses. Experimental Writers made more changes in their texts but did not produce revisions that were rated more effective than those of the Control Writers. Before discussing these findings, some cautions need to be pointed out. Clearly, a number of other analyses of these data must be performed. The measure of change reported here--the difference in overall length between Original and Revised texts--is an extremely crude one. Other measures of change that are more sensitive to the content and organization of the texts will be obtained before the conclusion that more change is made by Experimental subjects is completely accepted. Further, it is not just amount of change but also the nature of the changes that is important. and these too will be evaluated. Similarly, the measures of text effectiveness described so far are limited. The data in Tables 8 and 9 are based on global judgments by Raters who were trained on the procedural task and who carried out a number of ratings. Judgments by raters who were actually learning from the texts or performance data of subjects trying to use the texts to learn (as in the

video tapes we have for half of the Original texts) would be informative alternatives to the data in Tables 8 and 9. Similarly, more conceptually based evaluations of the content and organization of these texts can be carried out. Such further analysis may lead us to qualify these initial impressions of these data.

We also have a considerable amount of information about the properties of half of the original texts in the video-tapes of Readers thinking aloud while learning the procedures. We have not even begun to evaluate these data. They contain a wealth of information about the features of these texts that were easy and difficult to understand and how such features were used in the Experimental Writers' revisions.

The preliminary analyses we have carried out, along with our informal impressions from having watched subjects being run and having informally sampled from the Reader tapes, show that the Experimental Writers had available to them in these tapes much more information than they used. If this impression is correct, why did they not use the information? There are two sets of reasons. The first is motivational or emotional. We obtained three Readers for each Writer's texts to forestall the possibility that the Writer could dismiss any difficulties he observed as the fault of any particular Reader. To our surprise, even when they saw three different Readers having more or less the same difficulties, a number of the Writers <u>still</u> attributed the problem to the Readers and not to their own text. They would make comments like "Stupid readers" or "It's all there in the text." Indeed, the single most common revision among the Experimental Writers was to put a statement at the beginning of the text which asked the reader to read the instructions carefully! A second type of reason why the feedback was not more effective is that even when the Writer decided there were problems with his or her text they may not have known exactly what to do about them. New research we have planned will address many of these questions.

Another class of analyses we have been working on is the classification of the content of the descriptions. Using several sources as starting points, including our own intuitions about procedures and Miller's (1980) descriptions of the content of his data, we have developed a scheme for coding the content of our descriptions. Our current coding scheme is described in the Appendix B. Basically, this scheme classifies all of the content of each description into 10 categories of two broad types. The first type are those that have to do with direct statements of how to do the procedure, referred to by the heading of <u>Procedure</u> (see examples in the Appendix B). The second broad type are supplements to these direct statements, such as overviews, summaries, and examples (see the Appendix B). Since these all pertain to material whose aim is to help the reader, we refer to this set of categories by the general heading of <u>Guidance</u>.

We had a number of specific questions in mind as we developed these classes: How does the content of effective and ineffective descriptions differ? Do different procedural tasks elicit different types of descriptions? What type of content do Readers tend to have the most trouble with? What type of content is most likely to be revised by

Writers? What is least likely to be revised? Does this vary with procedural task or experimental condition? Do skilled writers revise differently than unskilled ones?

At the time this report is being prepared, only the first of these analyses has been completed. However, the results of this first analysis are quite provocative. For the Fix task where we had direct ratings of text effectiveness for each of the texts, we selected the four highest and the four lowest rated texts. We coded these texts using the scheme in the Appendix B. Several startling properties emerged, and are shown in the data summarized in Table 10. If we look at the overall proportion of text content devoted to direct description of the Procedure versus the content devoted to Guidance, we find a complete reversal between the best and the worst texts. Roughly two-thirds of the content of the best texts is devoted to Guidance and only a third to Procedure, while for the worst texts the proportions are reversed. Moreover, much of the difference is due to the presence of Examples. Roughly a third of the content of the best descriptions is taken up with examples, while only about ten percent of the worst ones are. Indeed, two of the four bad texts did not have any examples at all. This suggests that examples in particular and the kind of content coded by our Guidance categories in general may be important components of effective descriptions of procedures.

While these data suggest this, the conclusion is not firm. A number of issues need to be addressed, and these will be a major part of new research we hope to conduct. What are these issues? First, we have shown a correlation, but we do not know if the relationship is causal. Maybe our best texts were simply written by the smartest or most literate subjects. Second, how general is this finding? We need to examine other procedural tasks and other descriptive situations. Third, text effectiveness was measured by global ratings. Such ratings have a number of well-known limitations (Cooper, 1977). Thus, alternative ways of measuring text effectiveness need to be examined, both with our current data and with new data we hope to collect.

SUMMARY AND FUTURE PLANS

The program of research funded by this grant focussed on discovering the higher level strategies used by readers in comprehending simple texts. A variety of complex data have been collected in order to discover what these strategies are like. In addition, an initial, preliminary study has been carried out of how writers take into account the way in which their text will be understood by readers.

In a sense, the major purpose of the research has been to fill in details of the conceptual scheme sketched in the introduction to this report and described in greater detail in Olson et al. (1980, 1981). After three years of work the scheme appears to us as useful as it did at the outset. But now many of the concrete details of what readers are doing have been filled in.

Table 10

Properties of Best and Worst Rated Texts for Fix Procedure

	Subject	Mean Rating	Length	Content		
				Procedure	Guidance	(Examples)
Best Texts						
	2	6.1	518	.19	.80	(.35)
	4	5.8	455	.26	.71	(.63)
	15	5.8	554	.42	.58	(.27)
	21	5.8	515	.54	.42	(.12)
Mean		5.9	511	.35	.63	(.35)
Worst Texts						
	12	2.8	421	.89	.09	(0)
	16	2.2	428	.50	.50	(.24)
	22	3.0	145	.88	.12	(0)
	26	2.5	746	.62	.36	(.18)
Mean		2.6	435	.72	.27	(.11)

The work started during this grant is not yet completed. Several of the studies are still in progress, and the large data bases collected through the use of the thinking-out-loud methodology have not yet been completely analyzed. In turn, there are yet a number of journal articles to prepare. These activities will be pursued during the coming year (83-84).

The preliminary study of writing carried out towards the end of the grant represents the major new line of work to be pursued in the future. Already, grant proposals for new research on writing have been prepared and submitted to several agencies. Over the next few years a vigorous program of research on the nature of writing will be carried out.

TALKS AND PAPERS

Talks and Presentations

Olson, G.M. The process of story understanding. Talk presented at the Knowledge Representation Workshop, Brown University, June 1980.

Olson, G.M., Mack, R., & Duffy, S. Strategies for story understanding. Paper presented at the Annual Meeting of the Cognitive Science Society, Yale University, June 1980.

Olson, G.M., Duffy, S.A., Mack, R., Vincent, P., & Eaton, M. Cognitive processes during text understanding. Paper presented at the 22nd Annual Meeting of the Psychonomic Society, Philadelphia, November 1981.

Olson, G.M. Talk-aloud protocols as a method for studying text comprehension. Talk given at the Conference on "New Methods in the Study of Immediate Processes in Comprehension," University of Arizona, December 1981.

Duffy, S.A., Olson, G.M., Mack, R.L., Vincent, P.L., & Eaton, M. Readers' use of genre conventions in understanding and recalling academic essays. Paper presented at the Annual Meeting of the American Educational Research Association, New York, March 1982.

Olson, G.M., Duffy, S.A., Eaton, M., Vincent, P.L., & Mack, R.L. On-line question-asking as a component of story comprehension. Paper presented at the Annual Meeting of the American Educational Research Association, New York, March 1982.

Olson, G.M. Discussion. Paper presented as part of the Symposium on the Psychology of Questions at the Meeting of the American Psychological Association, Washington, D.C., August 1982.

Olson, G.M. Natural language descriptions of procedures. Talk given at Bell Laboratories, Murray Hill, New Jersey, February 1983.

Olson, G.M., Trahan, M., & Roshwalb, L. Natural language descriptions of procedures. Talk to be given at the Annual Meeting of the Psychonomic Society, San Diego, November 1983.

Publications (* indicates items included in Appendix C)

*Olson, G.M., Duffy, S.A., & Mack, R.L. Applying knowledge of writing conventions to prose comprehension and composition. In W.J. McKeachie (Ed.), Learning, cognition, and college teaching. San Francisco: Jossey-Bass, 1980.

*Olson, G.M., Mack, R., & Duffy, S. Cognitive aspects of genre. Poetics, 1981, 10, 283-315.

Olson. G.M. On language understanding process (in Chinese). Shinli Kexue Tonxuin. 1981. 3. 70-74.

*Olson. G.M.. Duffy. S.A.. & Mack. R.L. Thinking-out-loud as a method for studying real-time comprehension processes. In D. Kieras & M. Just (Eds.). New methods in the study of immediate processes in comprehension. Hillsdale. New Jersey: Lawrence Erlbaum Associates. 1984.

*Olson. G.M.. Duffy. S.A.. & Mack. R.L. Question-asking as a component of text comprehension. In A. Graesser & J. Black (Eds.). The psychology of questions. In press.

*Duffy. S.A. The role of expectations in sentence integration. Ms. submitted for publication.

Olson. G.M.. Duffy. S.A.. Eaton. M.E., Vincent. P.. & Mack. R.L. On-line question-asking as a component of story comprehension. Ms. in preparation.

Olson, G.M.. Trahan. M.. & Roshwalb. L. The composition and revision of natural language descriptions of simple procedures. Ms. in preparation.

Publication Plans

Several other uncompleted portions of this research are likely to lead to publications. The detailed analyses of the thinking-out-loud data using the scheme in Appendix A should produce a manuscript (possible of monograph length) on the detailed strategies of readers. The experimental work on controlled reading times will be published if the results of the current study (and any planned follow-ups) warrants. Indeed. we suspect that the rich data we have on hand for both comprehension and composition may lead to a number of other presentations and publications beyond those currently planned. Copies of all subsequent publications that result from the project will be forwarded to NIE.

REFERENCES

Anderson, R.C., & Biddle, W.B. On asking people questions about what they are reading. In G. Bower (Ed.), The psychology of learning and motivation (Vol. 9). New York: Academic Press, 1975.

Duffy, S.A. Predictive processing in story comprehension. Unpublished doctoral dissertation, University of Michigan, 1983.

Duffy, S.A., Olson, G.M., Mack, R.L., Vincent, P.L., & Eaton, M. Readers' use of genre conventions in understanding and recalling academic essays. Paper presented at the Annual Meeting of the American Educational Research Association, New York, March 1982.

Frase, L.T. Prose processing. In G. Bower (Ed.), The psychology of learning and motivation (Vol. 9). New York: Academic Press, 1975.

Kintsch, W. The representation of meaning in memory. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Knuth, D. The art of programming: Fundamental algorithms (Vol. 1, 2nd. ed.). Reading, Mass.: Addison-Wesley, 1973.

Meyer, B. The organization of prose and its effect on memory. Amsterdam: North-Holland, 1975.

Mifler, L.A. Natural language programming: Styles, strategies, and contrasts. Research Report RC 8687, IBM Thomas J. Watson Research Center, Yorktown Heights, New York, December 1980.

Miyake, N., & Norman, D. To ask a question, one must know enough to know what is not known. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 357-364.

Newell, A., & Simon, H.A. Human problem solving. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Olson, G.M., Duffy, S.A., Eaton, M.E., Vincent, P., & Mack, R.L. On-line question-asking as a component of story comprehension. Manuscript in preparation.

Olson, G.M., Duffy, S.A., & Mack, R.L. Applying knowledge of writing conventions to prose comprehension and composition. In W.E. McKeachie (Ed.), Learning, cognition, and college teaching. San Francisco: Jossey-Bass, 1980.

Olson, G.M., Duffy, S.A., & Mack, R.L. Thinking-out-loud as a method for studying real-time comprehension processes. In D.E. Kieras & M.A. Just (Eds.), New methods in reading comprehension research. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1984.

Olson, G.M., Mack, R.L., & Duffy, S.A. Cognitive aspects of genre. Poetics, 1981, 10, 283-315.

Turner, A., & Greene, E. Construction and use of a propositional text base. JSAS Catalog of Selected Documents in Psychology, 1978, 3, 58. (Ms. No. 1713)

APPENDIX A

Scheme for Coding Think-Aloud Protocols

The format for coding a protocol according to this scheme will be:

OP code[Att code(seg code)]

The attention codes which are outside the parentheses and inside the brackets define the general category of context in-formation being focused on. The entire unit containing the attention code and segment code is called an Attention Unit. The operation codes which are outside the brackets define the cognitive operation carried out on the content within the brackets. If an object of an attention unit is a clause then the clause is characterized by an X and is defined in a separate attention unit. It is not necessarily going to be the case that all of the arguments shown for the attention codes will be filled. Only those that are minimally necessary should be included. Dummies can be used for necessary roles that are not mentioned. It is also possible for more than one OP Code to appear for a given segment of the protocol if it happens to be an especially complex one. Likewise more than one Att code can be used for each OP code.

Attention Codes

1. PAct (verb,agent,object,indirect object,instrument). Intentional actions of a character. They do not include habitual actions of a character, but rather specific actions that occur at one point in time in the story. This also includes verbs which summarize a set of Pacts (e.g., to plan, to organize).

e.g. Lentil played the harmonica for Colonel Carter.
PAct(play,Lentil,harmonica,Carter)

Actions which are negated are Pacts.

e.g., Lentil did not play his harmonica.

(Close miss: The musicians could not pucker = Pstate)

2. Ev (verb,agent,object). An event is anything that actually happens to a character or object as a result of a non-intentional internal cause (e.g.,blowing up) or an outside force (e.g. being blown away by the wind). Any noticeable change in the general conditions in the story should be coded as an event (e.g., weather changes, dawn or dusk). If a statement is ambiguous as to whether it should be an act or an event

(i.e., the intentionality of a character is difficult to discern), it should be classified as an Event.

e.g. The wind blew Lentil over.
Ev (blew over,wind,Lentil)

e.g., The mountain blew up.
Ev (blew up, mountain)

e.g., Dawn arrived.
Ev (dawn)

by convention, general statements in which it is impossible to tell whether a subject intended an action or an event are coded as Ev.

e.g., something will happen.
Pred[happen, something)]

3. MAct (verb,agent,object). A mental action or event such as thinking, deciding, learning, forgetting. This does not include habitual mental actions that a character constantly performs (which are classified as H-Macts).

e.g. The stranger learned about the town.
MAct (learned,stranger,townsfolk)

4. MState (agent,state). An emotion or belief that a character is experiencing. These are non-enduring mental states. Enduring emotions or beliefs are coded as Chars.

e.g.The stranger was disappointed.
MState (str,disappointed)

5. PState (agent,state). The physical state or condition of a character or an object. These are non-enduring traits, such as being hungry or cold (enduring physical traits are coded as Chars). They can also refer to an inanimate object and a temporary state that it may be in.

e.g. The stranger was cold.
PState (str,cold)

The boat was filled with people.
Pstate(boat,filled with people)

6. Poss (possessor;possessed). This category shows ownership or possession.

e.g. This is the woman's house
Poss (Wom,house)

7. Char (topic, characteristic). General enduring characteristics of both a character or an inanimate object. This does not include characteristic actions physical or mental (which are coded as H-Pact or H-Mact). A Char could include a period of a person's past e.g. Lentil lived in Europe. Segments in which it is ambiguous as to whether the attention code should be an H-Pact or an H-Mact are to be coded as a char. e.g., Lentil tries hard.

e.g. The house was clean.
Char (house, clean)

The boat was small.
Char (boat, small)

8. Goal (agent, goal). Statements which convey the need, desire, or motive of a character are not classified as Chars or Mstates. Rather they are classified as Goals. Most goal statements will contain an embedded clause which actually states what the goal is. This embedded clause is coded.

e.g., Lentil wanted to make music. Goal (Lentil, X) x, H-Pact (make, Lentil, music)

9. Loc (relation, located entity, location1, location2). Describes the location of an action, entity, or event involving specific characters, or other entities. That is, these are the locations of specific things, as distinguished from the general location of the story, which is an aspect of Set.

e.g. The boy was under the bed.
Loc (under, boy, bed, 0)

10. Set (property). General characteristics of the setting, such as the weather, general conditions that hold, the historical period, and the general location of the story.

Alto is in Ohio
Set (Alto, in Ohio)

11. Und (object). Some story content got mentioned with no attention category.

e.g. This is Lentil.
Und (Lentil)

12. Ti (time reference). Explicit reference to the time line of actions and events in the story. Reference to something happening now is not coded as Ti unless it is marking a change of condition from an earlier time (i.e., an explicit contrast with an earlier time).

Lentil met Carter after the parade.
Pact(met,Lentil,Carter),ti(after the parade)

If the "time phrase" conjoins two clauses, each clause is coded with its own Act code, and the "time phrase" is coded as Ti:

e.g., It will start raining as soon as he gets out of his car.
Pred[Ev(rain),Ti(as soon as X),X,Pact(get out of,car)]

Predictions are not coded as Ti unless they pinpoint the future act or event on the story timeline.

e.g., Next, Lentil will play his harmonica.
Pred[Pact(play, Lentil, Harmonica),Ti(next)]

(Close miss: Colonel Carter will meet Lentil sometime in the future - Ti is not coded here)

13. Sto (aspect,content). A residual category for attention to general characteristics of the story. Two kinds of statements are typically labeled Sto:

general statements about the story itself:
The story seems to be about Indians.
Sto(Indians)

more specific statements about story content where no other attention code fits:
We will meet Colonel Carter later on
Sto(Carter)

This second type of Sto is often confused with an Und and an Ev.
Confusion with Und: While the above statement at least has an implicit reference to the story as a whole (i.e., that we will meet Carter later in the story), the typical Und does not (e.g., "Here's Colonel Carter Und(Carter).
Confusion with Ev: While the verb "meet" is normally classified as an Ev, the subject of the verb is not a story character. Thus the "meeting" is not a story event and is not coded as an Ev.

14. Sty (aspect,content). Reference to the form of the language, purpose of the sentence, or the way it was written including vocabulary syntax, and author's style. No story content should be mentioned.

Selfish is an unusual word to chose.
Sty(unusual,selfish)

That sentence really doesn't say much.

Sty(doesn't say much)

15. H-Pact (verb,agent,object,indirect object,instrument)
H-Pacts are habitual actions of a character. It is something that the character is known to do routinely throughout the story, or is given as background about a character. If it is ambiguous whether a segment should be a H-Pact or a H-Mact then it should be coded as a char.

e.g. Lentil played his harmonica on the way to school.
H-Pact(played.Lentil.harmonica.on way to school)

16. H-Mact (verb,agent,object)
H-Macts are habitual mental actions of a character. This would include a habitual thought of a character or a mental action that a character performs throughout the story or is given as background information about him.

17. Perc (perception,character,object)
A character perceives something through one of his senses. These are not intentional actions such as looking or listening. but rather seeing or hearing.

e.g.. The 2 men saw the canoe.
Perc(saw.2 men,canoe)

18. Ident (object identifying,purpose of object)
When the identity of an object or a character is being attended to. These are always comments about anaphoric reference. e.g.. the referent of a pronoun:

Is "he" the first Indian?
Ident(he,first Indian)

The coreferentiality of two noun phrases
This is "the one thing" to make Lentil unhappy
Ident(the one thing)

I expect that's rain hopefully.
Ident(wet sound.rain)

What is the one thing to make Lentil happy.
Ident(what one thing.make Lentil happy)

Operational Codes

1. Que. The subject asks a question about the story. It can be implicit or explicit. e.g. "I wonder where Wensleydale is" A question is also any statement in which the subject asks for more information.

Certain indirect Que's are easily confused with Gcoms. In general, if the indirect question concerns a character or object that is introduced in the current sentence, it is probably a question. If it has been mentioned before, or if there is no reason to suspect that the subject is confused, it is a Gcom.

2. Rep. Something is repeated directly from the current sentence in the story.

3. T Ret. The subject repeats something that was stated earlier in the text. If there is no explicit reference to the earlier text, code it as T Ret[Und(0)]. (See also Gcom).

4. P Ret. The subject repeats something that he stated earlier in his protocol (that was not stated earlier in the text). Pret takes precedence over Pred or Inf and other similar operation codes. Pret indicates a protocol statement in which the subject does not generate a new prediction or inference but rather retrieves an old one. Thus the retrieval operation is coded.

5. Pred. The subject makes a prediction about what he expects to read about next or later on in the story. Pred can also be thought of as inferences stated in the future tense.

6. Inf. The subject infers something from the essay that is not stated. You need to look at the story to distinguish whether something is being inferred or not. Inferences which are stated in the future tense are classified as predictions.

7. GK. The subject states something that is general knowledge. e.g. "Ohio is a Midwestern state." "You need to know" is often a lead in for a GK.

8. Cob. The subject says something about his own behavior, thoughts, or feelings.

e.g., I think I read this story before.

9. NCom. These are negative comments that people make about the nature of the text, such as "This doesn't make sense". If it is ambiguous whether the segment is negative code it as a GCOM.

10. PCom. These are positive comments people make about the text.

11. GCom. General comments that are neither specifically positive or negative. These comments include indirect questions concerning a character or object mentioned earlier in the story (see Que) when it is clear the subject is not confused.

You'll have to know who the speech was for.

You'll have to know who Old Sneep is.

Notice that these Gcoms are characterized by a WH-question word following the "need to know" phrase. Statements taking the form "You need to know that" are usually followed by a fact or inference from the story. These are not Gcoms but Trets or Infs.

e.g., You have to know that Old Sneep sat and grumbled. = Tret

12. Conf. The subject confirms a prediction as being carried out in the story. e.g. "Just as I thought, the stranger was let in."

13. Disconf. The subject admits that a prediction that he made was wrong, or an inference drawn was incorrect. The prediction and inference do not have to have been explicitly stated.

14. Sum. The subject summarizes and combines ideas from earlier in the text, or it can summarize one sentence.

Hints on coding

frequently, much of what a subject says can be captured by a single Op or At code rather than a more complicated coding. for example:

to indicate the kind of person who would sit and grumble
Gcom[HPact(sit and grumble, Old Sneep)]
HPact captures underlined portion of statement

this sentence refers back to the fact that he can't sing.
Tret[Char(can't sing, Lentil)].
Tret captures underlined portion

You would have to know who Old Sneep is to understand the sentence
Gcom[ident(Old Sneep)]
Gcom captures the underlined portions.

APPENDIX B

Coding Scheme for Descriptions of Procedures

Categories Pertaining Primarily to the Logic of the PROCEDURE

1. Action [ACT]. The physical or mental actions performed as part of the procedure. By convention, this category is used for the full predicate. Negative acts are also ACTs. Examples: "Find the middle card." "Put your left hand on the leftmost card." "If the card below your left hand is less than the card below your right hand, don't move the cards." (In the examples, the underlined portion refers to the part coded as the category being described.)

2. Qualification [QUAL]. Qualifications placed on actions. The test on whether something is a QUAL rather than part of the ACT is whether the action has reasonable alternatives other than the one described. If so, it is a QUAL. Examples: "Put your left hand on the leftmost card" (right is a reasonable alternative to left, above, center, etc. are reasonable alternatives to on the leftmost card).

3. Condition [CON]. The conditions relevant to performing an act, using the typical if-then logic common to programming. The entire condition is classified as a CON. That is, the subparts are not classified as QUALs or other plausible categories. Examples: "If the card below your left hand is less than the card below your right hand, don't move the cards." "Continue shifting your hands to the right until your right hand is no longer on the top of a card." Note: Sometimes a writer will state a rule such as, "Rolls of one, two, three, four, score one, two, three or four points respectively." We code these as an implicit condition and action, so that "Rolls of one, two, three, four," is the condition, the remainder the action.

4. Initialization [INIT]. Statements which describe the materials or other conditions relevant to setting up to do a procedure. Examples: "You are given a deck of cards and something to be used as a marker."

5. Repeat or continue statement [REP]. These are statements which say at a general level to repeat or continue an action already described. If the steps are elaborated or repeated again, the entire text fragment that includes the elaboration or repetition is classified as a REP, without any further coding of the internal constituents. Examples: "You continue this procedure until the cards are in increasing order from left to right." "Continue steps D & E." "Perform steps A through G again." "If your left hand is on the last card on the left, go to step 4."

Categories Pertaining Primarily to the GUIDANCE of the Reader

6. Overview [OVER]. Statements that are overviews of the objectives, goals, or content of the procedure. General titles for a description are coded as OVER. Examples: "These are your instructions for a card

sorting procedure." "This procedure allows one to order any number of cards from the lowest to the highest number."

7. Summary [SUM]. Statements that summarize in a general way parts or all of a preceding description of a procedure. Examples:

8. Organizational Markers [ORG]. Explicit indicators of the sequence, organization, or structure of the parts of a procedure. Headings or labels for sections of the description and step numbers are examples of ORGs. Examples: "Now, shift each hand to the right." "To begin, cards are dealt out, face up, from left to right." "1) Odd number".

9. Examples [EX]. Examples, illustrations, analogies used to convey in an explicit fashion the steps of a procedure. Examples: "For instance if you are working with 9 cards you will be marking the 6th card." "We will use (7) seven, for example."

10. Cognitive Aid [COGAID]. These are statements inserted to help the reader. Specific types of statements include warnings, reminders, attention directions, tests for the reader to check on comprehension, etc. Examples: "Please read through the instructions once and then go step by step."

APPENDIX C

Publications Enclosed

A copy of each of the following publications is enclosed:

Olson, G.M., Duffy, S.A., & Mack, R.L. Applying knowledge of writing conventions to prose comprehension and composition. In W.J. McKeachie (Ed.), Learning, cognition, and college teaching. San Francisco: Jossey-Bass, 1980.

Olson, G.M., Mack, R., & Duffy, S. Cognitive aspects of genre. Poetics, 1981, 10, 283-315.

Olson, G.M., Duffy, S.A., & Mack, R.L. Thinking-out-loud as a method for studying real-time comprehension processes. In D. Kieras & M. Just (Eds.), New methods in the study of immediate processes in comprehension. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1984.

Olson, G.M., Duffy, S.A., & Mack, R.L. Question-asking as a component of text comprehension. In A. Graesser & J. Black (Eds.), The psychology of questions. In press.

Duffy, S.A. The role of expectations in sentence integration. Ms. submitted for publication.

Skilled writers employ the conventions of writing used in their field of study and skilled readers know these conventions well. Teaching writing may involve making these conventions more explicit to students.

Applying Knowledge of Writing Conventions to Prose Comprehension and Composition

Gary M. Olson
Susan A. Duffy
Robert L. Mack

"Reading maketh a full man, conference a ready man, and writing an exact man," wrote Sir Francis Bacon. Reading and writing have long been viewed as among the highest achievements of human culture, but literacy is essential to modern technological society, and deficiencies in reading and writing are considered major social problems. Despite the

In conducting research for this chapter, one or more of the authors received support from the following sources: a Research Career Development Award from the National Institute of Child Health and Human Development (HD 00169), predoctoral training grant from the National Institute of Mental Health (MH 14254), and a grant from the Horace H. Rackham School of Graduate Studies.

obvious importance of these skills, we still know very little about how the mind executes the complex tasks of understanding prose or composing it. The study of prose comprehension has only recently entered the mainstream of cognitive psychology, while the study of writing has not yet made it.

Recent research on the way in which human beings process printed prose has focused on what the reader remembers from the text—in particular, in how either the substance of the text or its form influence what a reader remembers. Much less attention has been given to what the reader is doing while reading. Only recently have some initial efforts been made to develop models of the reader's general strategies for reading a text (such as Collins, Brown, and Larkin, in press; Hayes and Simon, 1974; Kintsch and van Dijk, 1978; Kintsch and Vipond, 1978).

In this chapter we will present some components of a model of prose understanding. In particular, we will examine the knowledge that readers have about the conventions of prose composition and what effects this knowledge has on comprehension. We will also offer some thoughts on the implications of our analysis for the study of writing.

Communication, Conversation, and Convention

Most social interactions are governed by norms or conventions. Language use is no exception. Grice, in an influential set of lectures (1967, 1975) described some of the conventions that are important in language use. He used the participants in a conversation as his model. He noted that conversation is not random talk, that the participants in a conversation appear to be engaged in a cooperative venture, such as trying to exchange information or opinions with each other. Grice formalized this general point in what he called the "cooperative principle," which states that each participant both produces and comprehends each utterance in relation to whatever general purpose or direction holds for the conversation at that moment. The general principle can be broken down into several more specific maxims (we use Clark and Haviland's [1977] restatements of Grice's maxims): (1) *Quantity*—Make your contribution no more and no less informative than is required; (2) *Quality*—Say only that which you both believe and have adequate evidence for; (3) *Relation*—Be relevant; (4) *Manner*—Make your contribution easy to understand by avoiding ambiguity, obscurity, and prolixity. These maxims are useful for accounting for the way in which

meaning is conveyed by implication. Consider an example derived from Grice: Fred is standing alongside his car on a street and says to Mary, an approaching pedestrian, "I am out of gasoline." Mary replies, "There's a station around the corner." The implied meaning of Mary's remark is that the station is a gasoline station, that it is currently open (or at least she believes it to be), that it is within easy walking distance, and so on. In other words, the force of her remark goes beyond its literal meaning, and this force is interpretable in light of the "cooperative principle" and its specific maxims.

Written Communication

The idea of cooperation applies to all forms of linguistic communication. However, since written communication differs from speech in many important ways, we need to describe the special nature of the writer-reader relationship and develop a list of conventions specific to this case. Many scholars have noted that there are important differences between written and spoken communication (Hirsch, 1977; Pratt, 1977; Ricoeur, 1976; Rubin, in press; Schallert, Kleiman, and Rubin, 1977). Since there are many types of written and spoken communication (Rubin, in press), we need to be specific about what situations we are referring to. We will consider two prototypical situations, face-to-face conversation for spoken language and simple texts like stories, essays, and articles for written language. Table 1 lists some of the differences we have derived from the sources referred to above. Most of these are based on the fact that the composition of a text and its comprehension occur at different points in space and time. As a result, texts do not allow for interactional give and take, and thus must be composed with deliberateness and care. The difficulty people have learning to write is often attributed to these differences (see, for example, Hirsch, 1977).

With the differences listed in Table 1 in mind, we can turn to an analysis of the nature of written communication. Figure 1 presents a schematization of the situation. The starting point is the writer, who has in mind a complex network of propositions that represent the complete message that is to be communicated via the text. A major convention of linguistic communication is that much of the intended message is transmitted by implication. The writer expects the text to interact with the reader's interpretive skills and general knowledge to produce in the reader's mind something close to what is in the writer's

Table 1. Characteristics of Writing Compared to Speech

Characteristic	*Description*
Permanence	Writing persists through time, while speech is highly transient.
Detachment	Both the content and form of written language is divorced from the immediate context in space and time.
Absence of feedback	Writing is a one-way process, without feedback.
Nonspecificity	Writing is typically addressed to a general audience rather than a specific individual.
Tellability	The topic of the text deserves the trouble involved in writing it up.
Organization	Written language is much more planned and organized than speech.
Formality	The language of writing tends to be more formal than speech.
Economy	Written language has less redundancy than spoken language.
Greater precision and detail	Written language can develop an idea, character, or setting in greater detail and with greater accuracy.
Greater complexity and abstractness of subject matter	Written language can develop much more complex ideas than spoken language.

mind. The writer's task is to select and organize a set of propositions to be included in the text that will be maximally effective in leading the reader to reconstruct the intended message. According to this model, there are fewer propositions in the text than there are in either the writer's or the reader's immediate representation of the text.

How does the reader go beyond the text in constructing a representation of its intended message? This is one of the most basic issues in the psychology of reading comprehension. There are two general types of knowledge the reader uses. The first is general knowledge about the nature of the world and the events, actions, and objects that populate it. Research in both cognitive psychology and in artificial intelligence has conclusively demonstrated that general knowledge plays an important role in understanding even simple sentences and texts. The second is knowledge about how textual transmission works. Writers select and organize propositions for inclusion in their text not only on the basis of what they think the readers know and what their specific purposes are, but in accordance with general principles of style and organization that

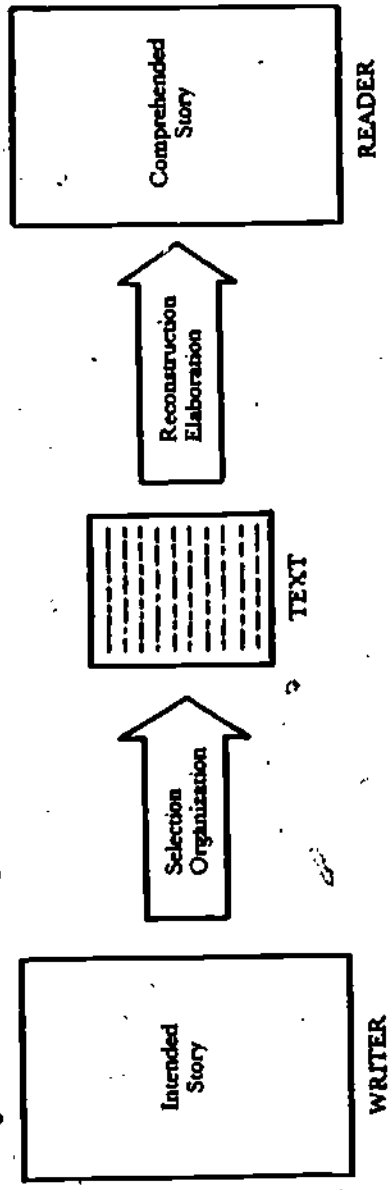


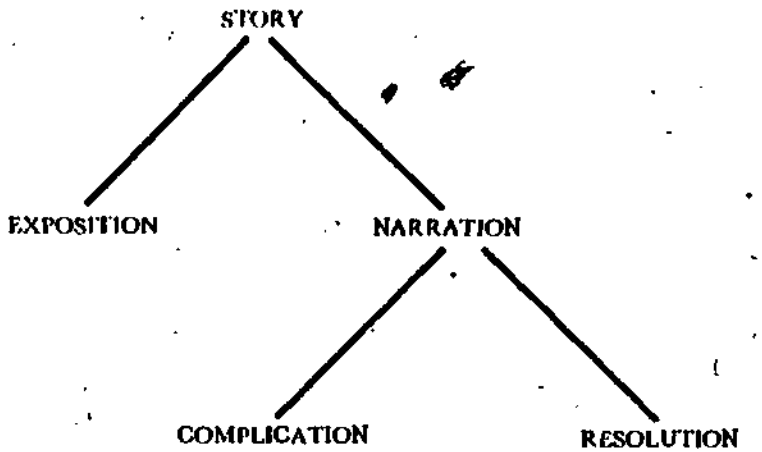
Figure 1. Schematic Representation of the Writer's and Reader's Tasks in the Writing Situation

apply to the genre, such as essay, research report, or story. The reader understands that the text was composed through a series of deliberate choices on the writer's part, and the reader understands the conventions that governed these choices. Further, the reader assumes that the writer had the reader's task in mind during composition, and was trying to make the reader's task (comprehension) possible. The reader interprets the sentences of the text in accordance with these general beliefs about the writer-reader relationship. In essence, the task of the writer is to guide the reader through a plot, an argument, or some other discourse structure. The reader expects this guidance. Accordingly, we have chosen to call the overriding principle for written communication the "guidance principle." Written communication is planned, one-way, and noninteractive, and the writer's role is to provide an appropriately orchestrated set of clues about the intended message. The reader assumes the writer is acting in good faith in the role of guide. An effectively written text is one in which the reader is in fact guided toward the reconstruction of the intended message.

Most of the recent research in cognitive psychology on text understanding has used simple stories as materials. Thus, we will use such stories to illustrate what we mean by conventions of comprehension and describe their effects on readers. Later we will describe comparable phenomena for two nonfiction genres, academic essays and magazine articles.

A simple story is a story with a single focus or plot that is told from a single point of view. Even simple stories have both underlying and surface levels of structure or organization. The underlying structure is an abstract representation of the information contained both implicitly and explicitly in the text of a story. Figure 2 presents a general representation of story structure. There is a network of background information, called the Exposition, and the core of the story (the Narration) consisting of Complication and Resolution. Story grammars of the kind developed by Mandler and Johnson (1977), Rumelhart (1975, 1978), and Thorndyke (1977) codify the complexities possible in the highly schematic structure shown in Figure 2. The text or surface structure of a story represents one embodiment of the underlying structure. The writer has many options in transforming an underlying structure into a text. First, the underlying structure is much more complete than the surface structure. Thus, different surface forms can vary in which propositions are selected. Among the factors that influence the selection are the writer's assessment of what the

Figure 2. The Major Elements of Story Structure



reader knows and the writer's purpose in telling the story. Second, the order of elements in the surface structure can vary. Presumably the underlying structure of a story has its propositions in their causal or temporal order. However, the writer can choose to present events in any order at all, as long as the underlying order can be reconstructed from the text. Similarly, though the Exposition is usually at the beginning, it can be delayed in a variety of ways (Sternberg, 1978). Third, specific versions of a story can vary in the point of view of the implied narrator. Scholes and Kellogg (1966) differentiate narrative from drama by noting that the former consists of both a story and a storyteller, the latter only a story. The storyteller can be one of the characters in the story, narrating in either first or third person, or he can be omniscient and uninvolved. Fourth, surface versions of the same story can vary in style, that is, in the selection of particular words, phrases, and sentence structures designed to create certain effects.

Let us now examine the conventions of composition shown in Table 2. Many of them are self-explanatory, so we will provide only a cursory description here (a more complete description will appear in Olson, Mack, and Duffy, forthcoming). Although these conventions apply specifically to simple stories, we believe they are similar to those for all simple texts, and we believe students can be aided both in their writing and in their reading by knowing them.

The first two conventions, *purpose* and *uniqueness*, are quite general, and apply to most forms of communication. They are based on

Table 2. Conventions of Composition for Simple Stories

Convention	*Description*
I. Purpose	The writer has a purpose in mind, such as to interest, amuse, or instruct.
II. Uniqueness	The story contains novel or unique elements, even if it follows a well-known formula.
III. Underlying Organization	The text is based on an underlying, abstract structure that is well formed for the particular genre.
A. Focus	There is one main line of development.
B. Overall Plan	The story has an overall plan that provides well-formed order to its individual propositions.
1. Coherence	All actions, events, or states are appropriately motivated, caused, or enabled.
2. Completeness	The structure leads to a resolution, with all loose ends tied together.
C. Representation	The world of the text is similar to, but intentionally not identical with, the world experienced by the reader.
IV. Surface Organization	The writer has a plan for presenting the story.
A. Omniscience	The writer knows all along how the story will end.
B. Point of View	The story will be told from one point of view.
C. Segmentation	The story is told by means of specific scenes or episodes.
D. Background	The writer will present sufficient background information so that the story can be understood.
E. Orderly Flow	The sequence of segments and propositions in the surface of the story is principled.
F. Connectivity	The writer will provide sufficient textual information so that the reader can make all the necessary inferential connections among the elements of the story.
G. Economy	Everything that is presented has a purpose within the surface plan of the writer.
H. Specificity	The plot of a story is presented by means of particular characters and events that are discrete in space and time.
I. Implication	The motivational or causal structure of the story is conveyed by implication.

Grice's principles that we described earlier. The books we examined on writing repeatedly pointed out how important it is to have a clear purpose in mind, and how often poor writing is characterized by the absence of clear purpose (for example, see Shaughnessy, 1977). According to uniqueness, even though the writer may be following a

highly stereotyped formula, there will be unique elements to the story. For instance, each murder mystery has its new twists.

The conventions of underlying organization describe the expectations a reader has about overall structure, and are based on the work on story grammars. The convention of *focus* is definitional for the simple stories we are considering. *Overall plan* states in an elementary way what the story grammars have tried to capture. The reader expects all of the elements of the story to fit into a coherent general framework, each element having its place in an overall plan and the entire plan having a clear resolution. The episodes of a story must be causally connected, but only those causally connected episodes in which conflict is created and resolved can be a story. The convention of *representation* has two parts. First, the reader expects the world of the story to be similar to the world we know. Even within highly stylized genres like science fiction there are strong expectations of correspondence to reality along with the obvious conventional departures from it (Scholes and Rabkin, 1977). Second, the reader also expects the world of the story to depart from reality in certain ways. The world of the story is an idealized world, or in Thornley's (1976) words, "an artistically disciplined representation of life" (p. 59). Thus, it is expected that stories will have characters who are stereotyped or larger than life, and that the sequence of events will often be highly improbable, with too many coincidences. Unlike the real world, the conflict in a story builds rapidly and clearly, and is totally resolved at the end. In short, the world of stories is a conventional world, similar to the real world in many important ways but also quite different from it.

The conventions of surface organization characterize the expectations a reader has about how the writer will tell the story. That of *omniscience* is a central one. The story is being told by someone who knows how it will end, and therefore the reader assumes that each element that appears is part of an orchestrated plan. This is the heart of the "guidance principle." *Point of view* is definitional for the simple genres we are considering. Only in complex short stories or in novels does one find a story being developed from multiple points of view. *Segmentation* states that the story will be told through a series of discrete units, in particular, through specific scenes or episodes that are separated in space or time. This is a property that narratives and dramas share. The tradition of scenes in a play is a good illustration of the point. The convention of *background* asserts that everything needed to make the story understandable will either be apparent on the basis of

general knowledge or will be provided by the writer. There are many conventions for presenting background information. A recent monograph by Sternberg (1978) describes and illustrates many of these. In very simple stories, the background is usually presented at the beginning. However, there is the well-known convention of *in medias res*, where the story begins in the midst of specific actions and the necessary background is woven into the development of the plot. This would seem to be a more difficult form for both the writer and the reader, but it is often used.

Orderly flow and *connectivity* are fundamental conventions of text processing, and have formed the heart of the theory developed by Kintsch (Kintsch and van Dijk, 1978; Kintsch and Vipond, 1978). Each element of the surface structure must be integrable into the developing network of propositions that represents the meaning of the text. The existence of well-known limits in immediate memory and attention suggest that propositions that cannot be immediately integrated will increase processing difficulty. Kintsch's model can be viewed as an explicit embodiment of these principles.

Economy is an extremely important convention for stories. Readers expect that everything in a story is there for a reason. If a small detail is mentioned, especially in isolation from other details, readers expect it may be important. Of course, in stories such as murder mysteries, attention to small details is elevated because of conventions associated with the genre. But we have seen much evidence in our research of readers paying special attention to details even where they are ultimately irrelevant. There is no conflict between this and the fact that details tend not to be well remembered (Mandler and Johnson, 1977; Meyer, 1975; Thorndyke, 1977). Some details may seem important when they are first encountered in the story and may receive special attention even if they are not central to the final representation. As a result, importance defined as a proposition's location in a story grammar hierarchy (see Thorndyke, 1977) may not covary with reading times in the same way that recall does.

Specificity is a closely related convention. Readers expect stories to be told by means of specific, concrete events and characters. The episodes that comprise the narration (see Figure 2) must occur at a specific time and place, and the characters must be particular. Statements about general patterns of events or the disproportional properties of characters are elements of the background. Sternberg (1978) has

described how the contrast between particular events and summaries of prior actions are used to differentiate narration from exposition. Our evidence indicates that readers are quite sensitive to this contrast.

Implication states that certain classes of information tend to be conveyed indirectly. In simple stories, the motive or causes of actions or events are almost never stated explicitly. The motivational structure of events in a story must be supplied by the reader. Interestingly, story grammars typically make this information explicit, supporting the view that they are best seen as descriptions of underlying structure.

Applications and Generalizations of the Conventions

In Olson, Mack, and Duffy (forthcoming) we describe in detail how the conventions in Table 2 relate to the strategies readers use in understanding simple stories. We have studied these strategies by examining a number of reader behaviors, especially the time taken to read each sentence in a text by subjects who are reading silently and the protocols provided by other readers who are asked to talk out loud while reading the text. We have studied the processing of both well-formed and ill-formed texts, and have verified that readers use knowledge of the conventions shown in Table 2 to guide their comprehension behavior. In reading a well-formed story, the reader generates hypotheses about the plot during the exposition, and then uses these—in conjunction with the conventions—to determine which sentences are most important and informative. When reading silently the reader devotes more time to these sentences, presumably to draw inferences and construct a coherent representation of the story. In ill-formed stories, which violate various of the conventions, comprehension is disrupted and readers are often misled or confused because their expectations about how the story ought to be told are violated. For instance, in Bartlett's (1932) famous "War of the Ghosts," readers have considerable difficulty constructing causal links between the individual episodes that follow each other in time (see Mandler and Johnson, 1977), and are unable to generate coherent hypotheses to guide their sentence by sentence comprehension of the text. By contrasting well-formed and ill-formed stories, we have been able to conclude that an important component of text readability is the extent to which general principles of the type shown in Table 2 are followed (see Olson, Mack, and Duffy, forthcoming).

Conventions of Nonfiction Forms

Educated readers (college subjects, for example) possess knowledge of the permissible underlying structures and the conventions of surface forms for simple stories, and this knowledge appears to influence their processing. Is this also true for other forms of writing such as the academic essay and the popular magazine article? On the basis of our examination of "how-to-write" books, the conventions of writing for these forms are clearly quite different from stories, but they appear to be just as well defined. Educated readers who know something about the conventions of composition for these forms use this knowledge during comprehension.

The Academic Essay. Both the academic essay and the magazine article are simple in the same way that the stories we have studied are simple: they have a single focus or line of development and they are written from a single point of view. The academic essay is the form most frequently encountered in books on rhetoric and in classes on expository writing. It is written in order to persuade the reader of the correctness of a thesis. The organization of the essay is rooted in the formal conventions of argumentation, and it is usually written with a thoughtful, serious reader in mind.

Books on rhetoric devote considerable attention to the underlying forms of argument an essay might employ. A deductive argument starts with an initial set of premises and supports the thesis through a series of intermediate deductions. Each step in the argument must be well formed according to the principles of logic. An inductive argument is one where the thesis is supported by a series of particular pieces of evidence, with each piece contributing general support for the thesis but none guaranteeing it. The underlying structure of an essay would be a canonical representation of the arguments developed in support of the thesis. Just as with a story, the elements of the argument can be embodied in more than one surface form. For example, a deductive argument has a natural order that starts with the initial premises and works its way toward the final conclusion via the intermediate deductions, but such an argument can be presented in various ways. The location of the conclusion or the thesis is one common source of variation: possibly at the outset, so the reader can have it in mind throughout the details of the argument; but possibly withheld until the end, particularly if the conclusion is unacceptable, controversial, surprising, or humorous. There are also conventions about which parts of the

deductive argument are made explicit, including the *enthymeme*, a syllogism whose major premise is implied, and the *sorites*, a chain of syllogisms in which only the final conclusion and the intermediate minor premises are stated explicitly. Most deductive arguments that appear in essays are enthymematic, and thus it is not surprising to find the enthymeme and the sorites discussed in books on rhetoric (see Brandt, 1970).

The strength of an inductive argument depends critically not only on the kind and amount of evidence but on its selection and ordering. That evidence which best supports the conclusion or is most representative of other evidence is clearly what the writer wants to select, and most authorities agree that the most persuasive ordering is to put the very strongest piece at the end. However, there are various ways to arrange the remaining pieces. The *climactic* order builds from the weakest evidence to the strongest, while the *Nestorian* order starts with the second strongest piece and then builds from the weakest to the strongest (Hughes and Duhamel, 1962).

The essay has much in common with the story, and the conventions in Table 2 can easily be modified to describe the principles of composition for essays. As in the story, the reader is using the text to try to extract the underlying structure, which in the case of an essay is an argument that focuses on a single thesis. The greatest differences would probably arise in the principles of surface organization. As one illustration, we pointed out that it is important for the Narration of a story to be told in specifics rather than generalities (specificity). This is because a story is about a specific set of characters interacting in specific locations at specific times. An essay, on the other hand, is written in a mixture of specific and general statements (Young, Becker, and Pike, 1970).

The Magazine Article. This much more heterogeneous class of prose forms than either the story or the essay is of interest because several authorities (Brandt, 1970; Dillon, 1977) observe that it is comprised of a blend of story and essay techniques. The magazine article is often written to inform or to persuade, but it must also entertain or interest the reader. The situations in which they are usually read are informal or casual, occasions when the reader does not want to engage in heavy intellectual work: in waiting rooms, on buses and planes, during lunch, in the bathroom, or relaxing on the sofa. As a result, the typical magazine article is much less formal than an essay, and the principles of composition are quite different.

Just like any other coherent text, an article has a beginning, a body, and a conclusion. But these parts are composed of a number of elements that are relatively unique to magazine writing (see Brandt, 1970; Dillon, 1977; Gunther, 1976). The article begins with a *lead*, a provocative or interesting anecdote or vignette designed to capture the reader's interest. The theme is usually stated shortly after the lead, but the relation of the theme to the rest of the article is not as tight as the relation of the thesis to an essay. The article often achieves *focus* by claiming to prove a point (Gunther, 1976), but the body or "proof" of the article at most illustrates or alludes to the point rather than presenting a logical argument for it. Brandt (1970), in characterizing reportorial writing, claims that the writer selects and organizes facts and generalizations in order to convey a perspective, but does not present a tightly structured argument. Coherence is achieved through "pseudo connections" between subordinate elements. This is the art of using various devices to create the illusion of the temporal connectedness of a narrative account or the logical development appropriate to demonstrating a thesis. One such device is to essentially list subtopics that could be relevant to the writer's point, as though the writer were going to develop the connections, but where in fact they are merely "called to mind" as illustrations of some point. Gunther (1976) describes similar devices, which he calls the "flash-by" and the "string-of-pearls" techniques in his influential practical guide on article writing. An article usually ends with another example or anecdote of the type used in the lead. Article writers are advised to achieve closure, to leave the reader feeling satisfied. But exactly what this means is harder to define for the article, which lacks the clear conflict resolution of the story or the logical coherence of the essay. Gunther (1976) sums up the approach to article writing by stating that the most important step is coming up with the right "slant" or "angle" for presenting the article content.

Articles in popular magazines often use elements of fictional writing. People, places, and events are sketched through anecdotes, vignettes, and quotes. The writer selects and organizes particulars in order to produce drama, surprise, suspense, or provocation. Generalizations are continually interwoven with narrative-like elements about people, and the motivations and reactions they exhibit. In other words, the writer simulates a narrative account that somehow conveys the point without presenting a carefully reasoned argument demonstrating it. Thus Dillon (1977) refers to the magazine article as a nonfiction story.

The "how-to" books on article writing give considerable attention to the conventions of this genre, and readers who have wide experience with magazine articles probably have some expectations about what they will find that are quite different than for stories or essays. Both of the latter forms have clear underlying structures, and the reader's main task is to extract these structures from the text the writer has provided. But we would not expect article readers to be trying to extract an underlying organization or adopt the hypothesis-testing mode that we have seen so clearly in our story research.

Conclusion

In this chapter we have presented an analysis of the relation between the writer and the reader of prose, and have used this analysis to describe some of the reader's strategies in comprehension. Most of our analysis has focused on the reader. However, we feel our work has important implications for the analysis of texts and for the writer. All of us share the intuition that texts vary in how easy they are to read. Many investigators have attempted to develop objective indices of text readability. Such indices would be useful in assessing the comprehension difficulties of readers and the composition difficulties of writers. Most readability indices, however, have been based on lexical or syntactic properties of text, and most have been only marginally useful (Kintsch and Vipond, 1978). Recently, Kintsch (Kintsch and van Dijk, 1978; Kintsch and Vipond, 1978) has proposed that the most useful indices of readability will be those that are based on a processing model of the reader's behavior. This sensible suggestion has some interesting implications. Perhaps the most important one is that readability becomes a joint function of the text and the reader, rather than being a function merely of the text. Some texts will be readable to almost everyone, and others for almost no one. But most texts will be differentially readable to individuals who vary in their specific substantive knowledge, knowledge of the conventions of various genres, particular reading skills, and general information-processing abilities. In support of this, Kintsch and Vipond (1978) have demonstrated how the readability of texts varies when they make different assumptions in their processing model about how much information (in the form of propositions) can be processed at one time, and how much can be held in immediate memory. Our research suggests that readability also varies with the writer's use of the conventions of writing. Thus our

work should contribute to the development of psychologically meaningful indices of readability.

The conventions listed in Table 2 are known by both the reader and the writer. Therefore, our analysis ought to have something to say about the process of writing. Up to now, there has been very little work done on the cognitive psychology of writing (some examples are Bruce and others, in press; Flower and Hayes, 1977). Most scholars divide the processes of writing into two broad categories, one associated with generating ideas and the other with putting the ideas into words. Our research on the principles of composition is relevant to the second category. However, what we need to do is to study the behavior of writers to see if we are really on the right track. Initially we might carefully examine the prose of writers who are at varying levels of proficiency, as well as examine successive drafts of a paper by reasonably skillful writers.

There is also an indirect way in which our analysis bears upon writing. It is widely believed that good reading and good writing go hand in hand. Good writers tend to do a lot of reading, and learning to read well seems to be an important component in learning to write well (Haynes, 1978). Our analysis is directed at making explicit the conventions of composition that are known to the skilled reader. If further research confirms our impression that this kind of knowledge is one component of effective reading, then explicit consideration of such knowledge in classes on writing might be useful for the learning writer. Further, a deeper understanding of the processes involved in prose comprehension and how these processes relate to composition might be a key to the crucial writing skill of self-criticism. The writer who can most successfully discover what is wrong with a particular draft of a manuscript ought to have the best chance of improving it in a revision. The principles we have proposed, and the techniques we have explored for identifying them, may prove to be useful tools for developing instructional methods aimed at acquiring these skills.

References

Bartlett, F. C. *Remembering: An Experimental and Social Study.* Cambridge: Cambridge University Press, 1932.

Brandt, W. J. *The Rhetoric of Argumentation.* Indianapolis: Bobbs-Merrill, 1970.

Bruce, B., and others. "A Cognitive Science Approach to Writing." In C. H. Frederiksen, M. F. Whiteman, and J. D. Dominic (Eds.). *Writing: The Nature, Development, and Teaching of Written Communication,* in press.

Clark, H. H., and Haviland, S. E. "Comprehension and the Given-New Contract." In R. Freedle (Ed.), *Discourse Production and Comprehension*. Vol. 1. Norwood, N.J.: Ablex, 1977.

Collins, A., Brown, J. S., and Larkin, K. M. "Inference in Text Understanding." In R. J. Spiro, B. C. Bruce, and W. F. Brewer (Eds.), *Theoretical Issues in Reading Comprehension*. Hillsdale, N.J.: Erlbaum, in press.

Dillon, M. T. *Magazine Article Writing*. Boston: The Writer, 1977.

Flower, L. S., and Hayes, J. R. "Problem-Solving Strategies and the Writing Process." *College English*, 1977, *39*, 449-461.

Grice, H. P. "Logic and Conversation." William James Lectures, Harvard University, 1967.

Grice, H. P. "Logic and Conversation." In P. Cole and J. L. Morgan (Eds.), *Syntax and Semantics: Speech Acts*. New York: Academic Press, 1975.

Gunther, M. *Writing the Modern Magazine Article*. (3rd ed.) Boston: The Writer, 1976.

Hayes, J. R., and Simon, H. A. "Understanding Written Problem Instructions." In L. W. Gregg (Ed.), *Knowledge and Cognition*. Potomac, Md.: Erlbaum, 1974.

Haynes, E. "Using Research in Preparing to Teach Writing." *English Journal*, 1978, *67*, 82-88.

Hirsch, E. D., Jr. *The Philosophy of Composition*. Chicago: University of Chicago Press, 1977.

Hughes, R. E., and Duhamel, P. A. *Rhetoric: Principles and Usage*. Englewood Cliffs, N.J.: Prentice-Hall, 1962.

Kintsch, W., and van Dijk, T. A. "Toward a Model of Text Comprehension and Production." *Psychological Review*, 1978, *85*, 363-394.

Kintsch, W., and Vipond, D. "Reading Comprehension and Readability in Educational Practice and Psychological Theory." In L. G. Nilsson (Ed.), *Memory: Processes and Problems*. Hilladale, N.J.: Erlbaum, 1978.

McCloskey, R. *Lentil*. New York: Scholastic Book Services, 1940.

Mandler, J. M., and Johnson, N. S. "Remembrance of Things Parsed: Story Structure and Recall." *Cognitive Psychology*, 1977, *9*, 111-151.

Meyer, B. *The Organization of Prose and Its Effect on Memory*. Amsterdam: North-Holland, 1975.

Olson, G. M., Mack, R. L., and Duffy, S. A. *Composition and Comprehension*. Manuscript in preparation, forthcoming.

Pratt, M. L. *Toward a Speech Act Theory of Literary Discourse*. Bloomington: Indiana University Press, 1977.

Ricoeur, P. *Interpretation Theory: Discourse and the Surplus of Meaning*. Fort Worth: Texas Christian University Press, 1976.

Rubin, A. D. "A Theoretical Taxonomy of the Differences Between Oral and Written Language." In R. J. Spiro, B. C. Bruce, and W. F. Brewer (Eds.), *Theoretical Issues in Reading Comprehension*. Hillsdale, N.J.: Erlbaum, in press.

Rumelhart, D. E. "Notes on a Schema for Stories." In D. G. Bobrow and A. Collins (Eds.), *Representation and Understanding: Studies in Cognitive Science*. New York: Academic Press, 1975.

Rumelhart, D. E. "Understanding and Summarizing Brief Stories." In D. LaBerge and J. Samuels (Eds.), *Basic Processes in Reading: Perception and Comprehension*. Hillsdale, N.J.: Erlbaum, 1978.

Schallert, D. L., Kleiman, G. M., and Rubin, A. D. *Analyses of Differences Between Written and Oral Language*. Technical Report No. 29. Urbana: University of Illinois, Center for the Study of Reading, April 1977.

Scholes, R., and Kellogg, R. *The Nature of Narrative*. New York: Oxford University Press, 1966.

Scholes, R., and Rabkin, E. S. *Science Fiction: History, Science, Vision*. New York: Oxford University Press, 1977.

Shaughnessy, M. P. *Errors and Expectations*. New York: Oxford University Press, 1977.

Sternberg, M. *Expositional Modes and Temporal Ordering in Fiction*. Baltimore: Johns Hopkins University Press, 1978.

Thorndyke, P. W. "Cognitive Structures in Comprehension and Memory of Narrative Discourse." *Cognitive Psychology*, 1977, *9*, 77-110.

Thornley, W. R. *Short Story Writing*. New York: Bantam, 1976.

Young, R., Becker, A., and Pike, K. *Rhetoric: Discovery and Change*. New York: Harcourt Brace Jovanovich, 1970.

Gary M. Olson is associate professor of psychology at the University of Michigan.

Susan A. Duffy is a doctoral candidate at the University of Michigan; B.A., Radcliffe College; M.Ed., Harvard. Her interest is in language comprehension processes.

Robert L. Mack is a doctoral candidate at the University of Michigan; B.A., Oakland University; M.A., Michigan State University. His interest is in language comprehension processes.

Poetics 10 (1981) 283–315
North Holland Publishing Company

COGNITIVE ASPECTS OF GENRE *

GARY M. OLSON, ROBERT L. MACK and SUSAN A. DUFFY

Readers and writers communicate via a text with the aid of a number of general and specific conventions. In this paper the general conventions governing written communication are described, and their application to two important genres, stories and essays, is analyzed. Data from subjects talking out loud while reading is used to obtain information about the knowledge and strategies readers employ while reading simple texts. Reading times collected from other subjects reading the same texts silently is used as converging evidence for evaluating the talking-out-loud data. A number of similarities and differences in the processing of stories and essays are reviewed. Story readers have an essentially prospective orientation, generating predictions and looking ahead to what is coming up. In contrast, readers of essays approach the sentence-by-sentence processing more retrospectively, fitting the current sentence in with earlier information that had been explicitly presented in the text.

It takes many types of knowledge to be a skillful reader. Recently, we have been examining how knowledge of the conventions of writing affects the comprehension activities of college readers (Olson *et al.* 1980; Olson *et al.* in preparation). Communication of information via written text is a specialized activity, and it is hardly surprising that a considerable amount of special knowledge would have to be acquired in order to be either an effective writer or reader. Our research has focussed on the types of knowledge skilled readers use while reading simple texts. In this paper we will examine the knowledge readers possess for two types of texts, namely, simple stories and essays of the type found in elementary rhetoric texts. Our basic claim is that skilled readers possess knowledge about the forms of these texts and their principles of composition, and that they use this knowledge during the process of understanding. We will describe the kind of knowledge college readers possess, and illustrate how they use it while reading. We will draw our data from a series of studies we have conducted using a variety of texts and tasks.

The article will be organized as follows. First, we will provide a general conceptual background for the type of analysis we have developed for text processing. As part of this background we will spell out our ideas on how communication via written texts work. Second, we will briefly describe the two genres we have investigated, simple stories and academic essays, presenting a set of defining characteristics for each. Third, we will describe the various tasks we have used to study the knowledge and processing strategies employed by skilled readers. Fourth, we will present a sample of our findings for the two genres in question. Finally, we will discuss the more general issue of knowledge of genre conventions as a component of what the skilled reader must learn.

* The work described here was made possible by a research grant from the National Institute of Education (NIE-G-79-0133) and by a Research Career Development Award from NICHD (5K04-HD00169) awarded to the senior author.

Requests for reprints or information should be addressed to: Gary M. Olson, Dept. of Psychology, University of Michigan, 330 Packard Road, Ann Arbor, MI 48104, USA.

0 304-422X/81/0000-0000/$02.50 © North-Holland

Conceptual background

Though we usually do not think of it in this way, communication via writing is a social process. Two people are interacting with each other, under the special condition that they are not at the same place at the same time. However, they are of necessity very much aware of each other. The effective writer must constantly have the reader in mind in order to produce a text that will have the desired effect. Similarly, the skilled reader must be aware of the author's intentions. Good writing represents a skillful selection and integration of material which, when combined with the reader's knowledge and strategies, leads to the outcome intended by the writer. Similarly, the reader, as part of what he or she needs to know in order to read the text, must understand both the writer's goals or intentions and the specific strategies used by the writer to select and arrange the material in the text. Both the writer and the reader understand that the text represents a specially selected and arranged set of propositions which will, in conjunction with the reader's knowledge, produce an approximation of the writer's intended message in the reader's mind.

In order to develop a model of how the writer and reader interact through a text, it is useful to consider some of the ways in which written communication differs from speech. While speech and writing have many similarities, and listeners and readers obviously employ many common processes and types of knowledge during comprehension, there are also some important differences that make the task of reading somewhat different. Many scholars have commented on these differences (e.g., Hirsch 1977; Pratt 1977; Ricoeur 1976; Schallert *et al.* 1977; Rubin 1980). Since there are many types of written and spoken communication (Rubin 1980), we need to be specific about what situations we are referring to. We will consider two prototypical situations: face-to-face conversation for spoken language and simple texts like stories or essays for written language. Table 1 lists some of the differences we have derived from the sources referred to above. Most of these are based on the fact that the composition of a text and its comprehension occur at different points in space and time. As a result, texts do not allow for interactional give and take and thus must be composed with deliberateness and care. A properly composed text is an organized, complete structure, worked out as a whole with an overall plan or purpose in mind. All of its parts serve some purpose. False starts, errors,

and other dysfluencies have been eliminated. Each sentence advances the text toward a point of closure, under the plan of the author. Relevant background information for the intended audience of the text is included if it cannot be inferred. The writer, if a good one, has taken some trouble to try to anticipate what the reader will need to know, and has included what would not be obvious. In short, the text must stand on its own. By the same token, because it is composed deliberately and carefully, it has a fluency not found in conversation. It is more formal. These many differences between speech and writing are major sources of difficulty for people learning to write (*e.g.*, Hirsch 1977).

Let us now turn our attention to an analysis of the nature of written communication. As noted before, there are certainly many similarities between reading and listening, and a number of the points we will make will be common to both types of comprehension. However, with the differences in table 1 in mind, we will also note some aspects of text processing that differentiate it from listening. Fig. 1 presents a schematization of the situation. The starting point is the writer, who has in mind a complex network of propositions that represent the complete message that is to be communicated via the text. A major convention of linguistic communication is that much of the intended message is transmitted by implication. The writer expects the text to interact with the reader's interpretive skills and general knowledge to produce in the reader's mind something close to what is in the writer's

Table 1
Characteristics of writing compared to speech.

Characteristic	Description
1. Permanence	Writing persists through time, while speech is highly transient.
2. Detachment	Both the content and form of written language is divorced from the immediate context in space and time.
3. Absence of feedback	Writing is a one-way process, without feedback.
4. Nonspecificity	Writing is typically addressed to a general audience rather than a specific individual.
5. Reflability	The topic of the text deserves to have the trouble taken to write it up.
6. Organization	Written language is much more planned and organized than speech.
7. Formality	The language of writing tends to be more formal than speech.
8. Economy	Written language has less redundancy than spoken language.
9. Greater precision and detail	Written language can develop a topic in greater detail.
10. Greater complexity and abstractness of subject matter	Written language can develop more complex ideas.

mind. The writer's task is to select and organize a set of propositions to be included in the text that will be maximally effective in leading the reader to reconstruct the intended message. The reader's task, obviously, is to use the text to reconstruct the writer's intended message. According to this model, there are fewer propositions in the text than there are in either the writer's or the reader's immediate representation of the text.

How does the reader go beyond the text in constructing a representation of its intended message? This is one of the most basic issues in the psychology of reading comprehension. In essence, the reader engages in a species of problem solving. The object of reading is to understand the intended message of the writer. The reader uses the text as data from which to generate hypotheses about what the text is about. The data from the text are evaluated with respect to knowledge the reader brings to the situation. This knowledge helps the reader to structure the text, though many texts also have explicit guidance in them about their structure. Inferences and elaborations are drawn from the data in the text. While actually reading the reader generates predictions about both the content and the structure of what is yet to come in the text.

There are two general types of knowledge the reader uses to engage in these activities. The first is general knowledge about the nature of the world and the conceptualizations about the world that have been formed as a result of prior experience. Many scholars have developed theories about the role of knowledge in processing external inputs, and have introduced concepts like schemata, frames, scripts, and MOPS (Bartlett 1932; Piaget 1952; Minsky 1975; Schank and Abelson 1977; Rumelhart and Ortony 1977; Schank 1979). Though there are many differences in detail among these concepts, there is no question that concepts like these are needed to account for language processing. Research in both cognitive psychology (*e.g.*, Bransford and Johnson 1973; Dooling and Christiaansen 1977; Bower *et al.* 1979) and in artificial intelligence (*e.g.*, Winograd 1972; Schank and Abelson 1977) has conclusively demonstrated that general knowledge plays an important role in understanding even simple sentences and texts. Because we know so much about the world we can understand elliptic references to what we know.

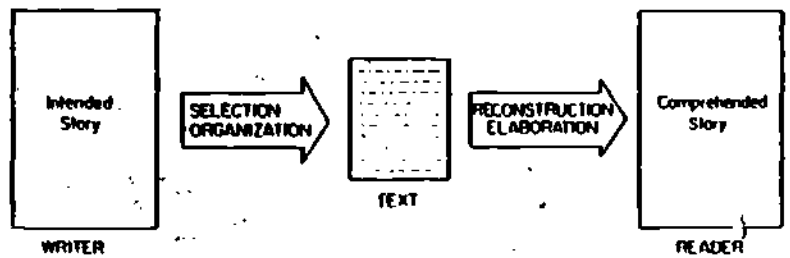


Fig. 1. Schematic representation of the writer's and reader's tasks in the writing situation.

The second type of knowledge used by the reader is about how textual transmission works. The writer selects and organizes propositions for inclusion in a text on the basis of what he thinks the reader knows and what his specific purposes are, in accordance with general principles of style and organization that apply to the genre. The reader understands that the text was composed through a series of deliberate choices on the writer's part, and the reader understands the conventions that governed these choices. Further, the reader assumes that the writer had the reader's task in mind during the composition, and was trying to make the reader's task (i.e., comprehension) possible. The reader interprets the sentences of the text in accordance with these general beliefs about the writer–reader relationship. In essence, the task of the writer is to guide the reader through a plot, an argument, or some other discourse structure. The reader expects this guidance. Accordingly, we have chosen to call the overriding principle for written communication the Guidance Principle. Written communication is planned, one-way, and non-interactive in real time, and the writer's role is to provide an appropriately orchestrated set of clues about the intended message. The reader assumes the writer is acting in good faith in the role of guide. An effectively written text is one in which the reader is in fact guided toward the reconstruction of the intended message.

Principles of text composition

We now want to describe a set of principles of composition that appear to be understood by writers and readers of simple stories. These principles are intended to describe the general properties of simple texts of the kind we have studied in our laboratory. Specifically, we have examined simple stories and academic essays. For technical reasons, we have confined our attention to stories and essays that are no more than 60 sentences long. This means that all the stories we have looked at are drawn from the fables, folktales, and children's stories that are typically investigated by researchers interested in story processing. They all have a single focus or plot that is told from a single point of view. Similarly, the academic essays we have used were taken from books on rhetoric. They all attempt to prove a thesis, make a comparison, or otherwise illustrate a single point. We feel that many elements of our analysis will generalize to a wider range of texts and text types, and will return to this point at the end of this article.

The principles we are about to present are intended to capture the general conventions about the composition of simple texts that are part of what a knowledgeable writer and reader knows. Though we initially developed principles like these for the specific genres of stories and essays, we found enough commonality to be able to construct this more general list. Later we will illustrate how these general principles apply to each of these two genres. Our central claim in this paper is that knowledge of these principles plays a role in comprehension. Presumably it also plays a role in composition, though we are only just beginning work on the writer.

It is likely that the principles about to be described also play a role in learning the specific conventions of a new genre, a situation we have not examined at all.

The principles have been developed on the basis of two broad classes of evidence. First, we surveyed the literature on story and essay writing, examining the advice given to writers about how to put a text together. We have also read monographs on the structure of various prose forms. Second, we have collected several kinds of data for both stories and essays. Some subjects read these texts sentence-by-sentence and talked out loud, stating any inferences, elaborations, predictions, interconnections, or other comments they felt compelled to make. Since the texts we used included both good and bad ones, we obtained much useful information about what readers expect in stories and essays and how this affects their processing. Other subjects read these same texts sentence-by-sentence on a computer screen and we timed their silent reading. We have also collected recall data and ratings of propositional importance for each text. All of these tasks will be described in more detail later.

Prior to presenting the list of principles it is necessary to make explicit a distinction that was implicit in our earlier discussion. A text has both an underlying structure and a surface structure. The underlying structure is an abstract representation of the information contained both implicitly and explicitly in the text. Each text type or genre has a set of principles describing acceptable underlying structures. For example, for simple stories, grammars of the type described by Rumelhart (1975, 1978), Thorndyke (1977), and Mandler and Johnson (1977) are best viewed as descriptions of possible underlying structures (Johnson and Mandler 1980, have specifically developed this view). Later we will describe possible underlying structures for academic essays.

The underlying structure of a text is an abstract description of its essential form. The surface structure represents one embodiment of the underlying structure. The writer has many options in transforming an underlying structure into a surface text. In fact, a given underlying structure can be transformed into many different surface texts. These surface texts can vary in a number of ways. First, they can vary in the way in which propositions are made explicit. The underlying structure is much more complete than the surface structure. Thus, different surface forms can vary in which propositions are selected. Among the factors that influence the selection are the writer's assessment of what the reader knows and the writer's purpose in writing the text. Second, the order of elements in the surface structure can vary. While the underlying structure of a text may have a specific causal, temporal, or logical order, the writer can choose to present the propositions of the text in any order at all, as long as the underlying order can be reconstructed from the text. Third, specific versions of a text can vary in style, that is, in the selection of particular words, phrases, and sentence structures designed to create specific effects. Fourth, certain texts permit variation in other factors. For instance, stories can vary in the point of view of the implied narrator. Scholes and Kellogg (1966) differentiate *narrative* from *drama* by noting that the former consists of both a story and a story-teller, while

the latter consists only of a story. The story-teller can be one of the characters in the story, narrating in either first or third-person, or he can be omniscient and uninvolved.

Let us now examine the set of principles shown in table 2. Many of them are self-explanatory, so we will provide only a cursory description here. Keep in mind that they apply specifically to simple texts of the type we have been studying. Later we will discuss how they translate into the two genres we have been investigating, simple stories and academic essays.

The first two principles, *Purpose* and *Balance of novel and familiar elements*, are quite general, and apply to most forms of communication. Books on writing repeatedly point out how important it is for the writer to have a clear purpose in mind, and how often poor writing is characterized by the absence of a clear purpose (e.g., Shaughnessy 1977). According to *Balance of novel and familiar elements*, even though the writer may be following a highly stereotyped formula, there will be unique elements. But the newness must be blended with familiar elements. A text works by building the new upon the foundation of the old. Later we will demonstrate the importance of these general principles for the comprehension of simple texts.

The principles of underlying organization describe the expectations a reader has about overall structure. The writer is trying to embody the underlying structure in an effective surface text. The reader is mainly trying to reconstruct the underlying structure. The principle of *Focus* is definitional for the simple texts we are considering. *Overall plan* states in an elementary way the essence of what we mean by underlying structure. The reader expects all of the elements of a text to fit into a coherent general framework, each element having its place in an overall plan and the entire plan having a closed or optimal structure. The principle of *Conventional world* has two parts. First, it is expected that the world of the text will be similar to the world we know. Even within highly stylized narrative genres like science fiction or highly abstract essays there are still strong expectations of some degree of correspondence to the world we know from experience. However, it is also expected that there will be systematic, conventional departures from the world of our experience. For instance, the world of stories is an idealized world, or in Thornley's words, "an artistically disciplined representation of life" (1976: 59). Thus, it is expected that stories will have characters who are stereotyped or larger than life, and that the sequence of events will often be highly improbable, with too many coincidences. Unlike the real world, the conflict in a story builds rapidly and clearly, and is totally resolved at the end. In short, the world of stories is a conventional world, similar to the real world in many important ways but also quite different from it. Similarly, with essays, a bit of abstraction and idealization is involved in putting together an argument or defending some general thesis. Yet it is essential that there be correspondences to the real world, or else the argument will be devoid of meaning.

The principles of surface organization characterize the conventions that govern

Table 2
Conventions of composition for simple texts.

Convention	Description
1. Purpose	The writer has one or more specific purposes in mind.
2. Balance of novel and familiar elements	The text contains a balance of familiar and novel elements.
A. Familiarity	The text makes contact with things the reader knows about.
B. Novelty	The text contains new or distinctive elements.
3. Underlying organization	The text is based on an underlying structure that is appropriate and well-formed for the particular genre.
A. Focus	There is one main line of development.
B. Overall plan	There is an overall plan that provides a well-formed organization for the propositions of the text.
(i) Coherence	Each individual proposition fits in a well-ordered way into the general plan.
(ii) Completeness	The overall organization has a closed or optimal structure.
C. Conventional world	The text is based on a world that is similar but not identical to the real world.
4. Surface organization	The writer has a surface plan for presenting the surface propositions of the text.
A. Omniscience	At any point in the text the writer knows where it is going.
B. Audience	The text is written with an audience in mind.
C. Scaffolding	The writer presents enough supporting material in the text so that the reader has sufficient background to be able to reconstruct propositions where necessary.
D. Segmentation	The text is organized into surface chunks.
E. Connectivity	The writer provides sufficient textual information so that the reader can make all the necessary inferential connections among the elements of the text.
F. Economy	Everything that is presented has a purpose within the surface plan.
G. Orderly flow	The sequence of segments and propositions in the surface of the text is principled.
H. Language	There are conventions for the surface language of each genre.
(i) Signalling	Appropriate surface signals will be used to mark transitions, etc.
(ii) Level	Appropriate levels of language will be used.
I. Genre-specific conventions	In addition to the general conventions listed above particular genres will often have specific text conventions governing the surface form.

the transformation of the underlying structure into an actual text. As can be seen from table 1, there are a number of general principles of surface structure but there are also idiosyncratic ones that apply to particular genres. The present discussion will focus on the common ones. Some idiosyncratic ones will be mentioned later when we discuss the particular genres we have investigated. The principle of *Coherence* is a central one. The text is being presented by someone (*i.e.*, the writer) who knows where it is going, how it will end. Therefore the reader is able to assume that each element of the surface structure that appears is part of an orchestrated plan. This has enormous implications for comprehension, as we will see later. It is the heart of the Guidance Principle. *Audience* characterizes the author's deliberateness in directing the text toward some specific group of readers who have particular characteristics. *Scaffolding* indicates that the writer is aware of the reader's need to have a superstructure constructed around which the elements of the text can be assembled. *Segmentation* states that the text will be written through a series of discrete surface chunks. *Connectivity* is closely related to *Scaffolding*, in that it is expected that all the links between the elements of a surface structure are constructable, either from direct evidence provided in the text or from a combination of surface hints and prior knowledge. Closely linked to this, however, is the principle of *Economy*, which states that there is a contrasting pressure to make the surface form as economical as possible. This leads the reader to assume that everything encountered in the surface text is there for a reason. This does not mean that everything is of equal importance, or will be equally well remembered. But each element has a purpose in the unfolding text. It is there for a reason, albeit perhaps a transient one. *Orderly flow*, along with *Connectivity*, are at the heart of the text processing theory developed by Kintsch (Kintsch and Vipond 1978; Kintsch and van Dijk 1978; Miller and Kintsch 1980). Each element of the surface text must be integratable into the developing network of propositions that represents the meaning of the text. The existence of well-known limits in immediate memory and attention suggest that propositions that cannot be immediately integrated will increase processing difficulty. Kintsch's model is an explicit embodiment of these principles. Finally, *Language* indicates that there are specific conventions governing the type of language that is appropriate for performing various functions within a genre. For instance, the events of a story are expected to be told through specific, concrete events and characters, while background information is given in more generic language. The argument in an essay unfolds through an appropriate blend of abstract and concrete language. We will see later that our evidence indicates that readers are quite sensitive to shifts of language in a text.

Simple stories and essays

Let us now turn our attention to the two genres we have been investigating in our research. In this section we will describe the properties of these genres, and will

relate them to the general principles of composition described earlier in table 2. As we shall see in the next section, college readers use knowledge of the properties of these genres during the process of understanding. In particular, they know something about the types of underlying structures allowable for the genre and in addition something about the conventions for transforming an allowable underlying structure into an acceptable surface text. Knowing these things, they are able to work backwards from the elements of the surface text they are presented to a representation of the underlying structure conveyed by the text. Clearly, strategies based on this kind of knowledge will go astray on those occasions when a text is ill-formed in either its surface or its underlying structure.

Stories

The fables, folktales, and children's stories that have been investigated intensively by cognitive psychologists in the past few years have a simple and straightforward structure. Fig. 2 presents in highly schematic form the basic pattern of the underlying structure of these stories. There is a network of background information, called the Exposition, and the core of the story (the Narration) consisting of Complication and Resolution. In essence, a story consists of a problem or complication which gets resolved by the intentional actions of one or more central characters. Of course, even in simple stories there are many ways in which the simple schema shown in fig. 2 can be made more complex. Further, there are constraints on what kinds of complications and resolutions will produce acceptable stories (Brewer and Lichtenstein 1980; Wilensky 1980; Olson in prep.). But the general pattern of complication and resolution is at the heart of the types of simple stories we and others have investigated.

The conventions of composition for simple texts can be translated quite directly for these stories. The purpose of such stories is typically to amuse or entertain, though of course many are also used to instruct. These stories achieve a balance between the familiar and the novel by following certain stereotypic patterns of plot development with familiar character types, but doing so with particular novel twists or surprises of detail. Later we will see quite clearly the consequences of this for

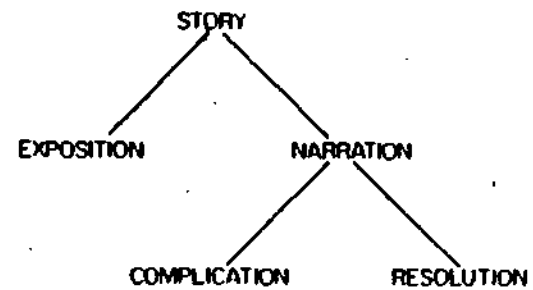


Fig. 2. The major elements of story structure.

processing. At the level of underlying structure the story follows a basic plan of conflict and conflict resolution, with a number of acceptable forms of complication allowed within this general framework. The reader expects all of the elements of the story to fit into a coherent general framework, each element having its place in an overall plan and the entire plan having a clear resolution. The episodes of a story must be causally connected, but only those causally connected episodes in which conflict is created and resolved can be a story.

The principles of surface organization characterize the way in which an underlying story gets told. In this regard, the principle of *Omniscience* is a central one. The story is being told by someone who knows how it will end, and therefore the reader assumes that each element that appears is part of an orchestrated plan. This is the heart of the Guidance Principle. The *Scaffolding* principle asserts that everything needed to make the story understandable will either be present on the basis of general knowledge or will be provided by the writer. In stories, this information is frequently in the form of background or setting information. There are many conventions for presenting background information. A recent monograph by Sternberg (1978) describes and illustrates many of these. In very simple stories, the background is usually presented at the beginning. However, there is the well-known convention of *in medias res*, where the story begins in the midst of specific actions and the necessary background is woven into the development of the plot. Sternberg describes other techniques as well. Though we have not investigated this directly, we could guess that the placing of background information in locations other than the beginning could complicate the task for the reader, though the writer can do much to minimize the difficulty by the way in which the information is introduced.

Segmentation for a story means that it is told through a series of discrete units, in particular, through specific scenes or episodes that are separated in space or time. This is a property that narratives and dramas share. The conventional organization of a play into scenes is a good illustration of the point.

Orderly flow and *Connectivity* are fundamental principles of text processing. Each element of the surface structure must be integratable into the developing network of propositions that represents the underlying story. *Economy* is an extremely important principle for stories. Readers expect that everything in a story is there for a reason. If a small detail is mentioned, especially in isolation from other details, readers expect it may be important. Of course, in stories such as murder mysteries, attention to small details may be elevated because of conventions associated with this specific genre. But we have seen much evidence in our research of readers paying special attention to details even where they are ultimately irrelevant. There is no conflict between this and the fact that details tend not to be well-remembered (Thorndyke 1977; Mandler and Johnson 1977; Meyer 1975). Some details may seem important when they are first encountered in the story and may receive special attention even if they are not central to the final representation. As a result, importance defined as a proposition's location in the hierarchy of a theory of story structure (*e.g.*, Thorndyke 1977) may not covary with reading times in the same way that recall does (but see, *e.g.*, Cirilo and Foss 1980).

Readers have specific expectations about the *Language* in which stories will be told. In part, they expect them to be told by means of specific, concrete events and characters. The episodes that comprise the heart of the story must occur at a specific time and place, and there must be particular characters involved in the episodes. Statements about general patterns of events or the dispositional properties of characters are elements of the background. Sternberg (1978) described how the contrast between particular events and summaries of prior actions is used to differentiate Narration from Exposition. Our evidence indicates readers are quite sensitive to this.

We have found at least one nice example of a genre-specific convention for stories. The motives of characters or causes for action are almost never stated explicitly. The motivational structure is almost always implicit, though it is clearly important for understanding the causal connections among events in the story. The task of figuring out why a character has done something is complex, and requires a general theory of goals and plans on the part of the understanding system (*e.g.*, Wilensky 1978, 1980). No doubt there are many other such conventions, but this particular one has very important consequences for processing. The implicit information must be recovered in order to understand fully the connections among the events.

Simple essays

Unlike the simple stories we have studied, simple essays do not have as well-defined a form at either the underlying or surface level. The simple story has a tightly organized structure, organized around a plot, a series of events that are causally related and that unfold in the complication-resolution plan described earlier. However, essays can have many forms because they have many purposes. Further, each type of essay seems to be governed by a looser set of conventions than those found for stories. This is not surprising, since even simple essays are written for a variety of purposes and the mapping of purposes onto rhetorical strategies is quite flexible. However, though this genre is more loosely structured than the simple story, readers do have expectations about the types of structures they will find. Let's look at some simple examples.

One common type of simple essay is the inductive argument. The author's goal is to convince the reader of a thesis, and evidence relevant to the thesis is presented as the heart of the essay. At the level of underlying structure, inductive arguments do not have a linear organization analogous to the causal, temporal organization found in stories. There is the thesis itself and statements of the evidence pertinent to the thesis. Any particular thesis has the potential for an indefinite number of evidence statements that might be relevant to it. However, as a practical matter of text convention, a thesis is typically supported by a modest number of evidence statements. In the types of simple texts we have examined, usually no more than two or three distinct types of evidence are cited. Rhetoric books often describe various princi-

ples for selecting evidence (see, e.g., Baker 1976; Hughes and Duhamel 1962; Kinneavy 1971; Payne 1969).

At the surface level, there are different ways to arrange an inductive argument. Following the principle of *Scaffolding*, however, the writer must always give the reader enough introductory information to make the unfolding argument structure apparent. Typically, this is best done by presenting the thesis at the beginning of the essay. The evidence is then presented, and the thesis is restated at the conclusion. Alternatively, the evidence can be presented first, leading to the statement of the thesis at the end. If this is done, however, the writer must take special care to inform the reader that the evidence supports a thesis which will not be revealed until the end of the essay. The evidence itself can be ordered in different ways, embodying the principle of *Orderly flow* for essays. These different orderings reflect the persuasive purpose of the inductive essay. Most authorities agree that the most persuasive ordering is to put the very strongest piece of evidence at the end. However, there are various ways to organize the remaining pieces. The *climactic* order builds from the weakest evidence to the strongest, while the *Nestorian* order starts with the second strongest piece and then builds from the weakest to the strongest (Hughes and Duhamel 1962). Unlike stories, *Segmentation* for the essay takes the form of paragraphs with information chunked according to topic. There are principles which govern sentence organization within the paragraph. For example, Kieras (1978) has studied readers' expectations about the placement of topic sentences within paragraphs. Our own data suggest that readers also have expectations about how the principle of *Language* is instantiated for essays. Readers expect a paragraph to be a blend of abstract statements of general principles and specific, concrete illustrations or examples.

Another common kind of essay is one which compares and contrasts two entities. For example, in work which we will not be describing here, we used two such essays, one in which a football halfback and a ballerina were compared, with emphasis on their similarities, and another in which two very similar maladies, heat exhaustion and sunstroke, were compared, with emphasis on their differences. There are two dominant types of organization for such essays: one in which all the properties of one entity are described before all those of the second, and another in which the two entities are described concurrently, with parallel properties described in alternation for the two. This contrast is not particularly surprising, perhaps, but it does influence what readers expect to find once they realize the purpose of the essay.

Empirical investigations of text processing

Let us now briefly describe the types of investigations we have conducted. Our strategy has been to use a number of different tasks to try to understand what it is that readers are doing while reading. In this section we will describe these tasks, and in the next section present some results from these tasks for stories and essays.

The texts

All of our texts are fairly simple and short so we can use them in different tasks without creating very long sessions. We used four different stories (the numbers in parentheses indicate the number of sentences per story):

Lentil (52): This story was taken from McCloskey (1940), and is a well-formed children's story. Since we will be discussing it in some detail later, a synopsis of its plot appears in table 3.

Stranger (44): This story was taken from Finlay (1969). It is also a well-formed story, a fable about the costs of greed and selfishness.

Circle Island (18): This story was originally introduced by Dawes (1966). The version we used was taken from Thorndyke (1977). Although this story fits Thorndyke's story grammar, it is not well-formed for reasons we will describe later.

War of the Ghosts (28): This classic was used by Bartlett (1932) in his studies of schemata. It is an American Indian folktale, and because it uses conventions and knowledge unfamiliar to most of us it is quite difficult to understand.

Our choice of these four stories was quite deliberate. We wanted two well-formed and two ill-formed stories so we could contrast our readers' behavior for these two types.

Table 3
Synopsis of "Lentil".

Story part	Content
Setting	A small midwestern town called Alto, Ohio
Characters	Lentil – a boy who wanted to sing and whistle, but couldn't so learned to play the harmonica instead
	Colonel Carter – a rich, important man who had given many buildings and parks to the town of Alto
	Old Sneep – an old, crabby fellow who complains about everything and everybody
Plot	Colonel Carter is returning to Alto after a two-year absence. The town decides to have a celebration. Old Sneep mutters that Colonel Carter needs to be taken down a peg or two. Preparations are made for the celebration: flags and signs are put up, the band is at the station, the mayor has a speech ready. The train arrives. As Colonel Carter steps from the train, Old Sneep is seen by all on the top of the station sucking a lemon. The band gets all puckered up and can't play. Everyone is silent and embarrassed, and Colonel Carter begins to look angry. Lentil saves the day by stepping up and playing on his harmonica. Colonel Carter is pleased, and a happy celebration is had by all, including Old Sneep.

Table 4
First two paragraphs of *Carpeting* essays.

Well formed	Ill-formed
I should start by admitting that as little as five years ago carpeted classrooms would rightly have seemed too fanciful and expensive.	As little as five years ago carpeted classrooms would rightly have seemed too fanciful and expensive.
The carpeting then available would have been costly, difficult to maintain, and would have required frequent replacement. Now, however, because of improved materials, the arguments in favor of extensive carpeting seem a great deal more plausible. New indoor–outdoor synthetics are stain resistant, fade resistant, durable and inexpensive. They have made carpeting seem much less a luxury than a reasonable, even desirable alternative to tile floors.	The carpeting then available would have been costly, difficult to maintain, and would have required frequent replacement. Improved materials have made the arguments in favor of extensive carpeting seem a great deal more plausible. New indoor–outdoor synthetics are stain resistant, fade resistant, durable and inexpensive. They have made carpeting seem much less a luxury than a reasonable, even desirable alternative to tile floors.
Briefly, there seem to be three central arguments in favor of carpeting.	
First, of course, carpeting is attractive. Now, admittedly, modern technology offers a great variety of attractively colored tiles. The days of drab, institutional grays, greens, and browns in tile are over. But while tile may approach carpeting in terms of color, it has a hard and unattractive texture. Carpeting, on the other hand, is colorful, attractive to the touch, and comfortable to walk on.	Carpeting is attractive. Modern technology offers a great variety of attractively colored tiles. The days of drab, institutional grays, greens, and browns in tile are over. While tile may approach carpeting in terms of color, it has a hard and unattractive texture. Carpeting is colorful, attractive to the touch, and comfortable to walk on.
It goes a long way toward creating a pleasant atmosphere all of us would like to work in, both in and out of class. Richly colored carpeting, such as bold reds often used in banks and commercial offices, would make our facilities less institutional. Bright carpeting can easily make attractive an area that would otherwise seem Spartan and sterile. In short, carpeting seems desirable simply because it is more attractive to look at and walk on than tile.	It goes a long way toward creating a pleasant atmosphere all of us would like to work in, both in and out of class. Richly colored carpeting, such as bold reds often used in banks and commercial offices, would make our facilities less institutional. Bright carpeting can easily make attractive an area that would otherwise seem Spartan and sterile.

We will also report data for two essays, each of which had two versions created by embodying the same underlying structure in two surface forms. The essays we used were:

Carpeting (37): This persuasive essay argues that carpeting is really more cost-effective than tile for classroom floors. The standard version was taken from Baker (1976). We created an ill-formed version by removing the signalling devices from the surface structure. The first two paragraphs of both versions of this essay are given in table 4.

Ice Age (22): This essay describes in a fairly factual way the possibility that we may be facing another ice age. It was taken from Jones and Faulkner (1968). The two versions differed in that one presented the thesis at the beginning followed by the evidence, while the second presented the evidence first, followed by the thesis. This second version is ill-formed because it does not make it clear to the reader that the evidence will precede the thesis statement.

For all of the texts used in these studies, an indication of the importance of each constituent was obtained by having an independent group of subjects cross out half the material in each text and then obtaining a profile of the frequency with which individual elements were deleted. This information will be referred to as *Importance Ratings* later in this paper.

Subjects

All of the subjects in the studies to be described were college students who were paid at a rate of $ 2.50 an hour for their participation.

Tasks

We used two different tasks with each of these sets of materials, using different groups of subjects for each task.

(1) Talking out loud

The subject was shown each sentence in the text one at a time, in sequence, and was asked to talk out loud into a tape recorder about a number of things. In particular, we asked subjects to talk about any inferences or elaborations they felt compelled to draw on the basis of the current sentence, any connections they saw between the current sentence and any prior ones, any predictions about what might be coming up, and any comments they had about the text or their understanding of it. Subjects who talked out loud to the essays were also asked to comment on the role an individual sentence had in the overall organization of the essay if they felt compelled to. The tapes from these sessions were transcribed, and checked, and the transcriptions segmented into 'idea' units and the units classified by type of statement. The segmentation and classification were checked for reliability by having two people perform these tasks.

(2) Reading time

Subjects were presented with each text at a computer terminal. Each time they pressed a key the next sentence of the text appeared. Only one sentence was shown

at a time. Subjects were told to read the text as normally as possible. Those reading the stories were told that later on we would explore how well they understood each story. Essay readers were told that they would later be asked to write a one-sentence summary of each essay. The primary data from this task are the times subjects devoted to each sentence in each text.

Processing simple stories

What is a reader doing while reading a simple story? This question has been at the heart of a series of investigations we have been carrying out. In this section we will present a broad sketch of our view of what the reader is doing, and then present several sets of data from our research that provides support of this view.

The reader of a simple story is confronted with the following task. A text embodying various aspects of the story is available as input. The text will be processed against a background of several types of knowledge that are relevant to it. The final product of understanding will be an interpreted representation of the essential elements of the story. However, with most stories, this representation must be constructed from incomplete information in the text. Much that is important to interpreting the elements of the text as a story is left implicit.

To try to give a concrete idea of what readers are doing while reading a simple story, let us describe some of our talking-out-loud data for several stories we have been examining. Table 5 presents the relative frequencies of different types of things readers talk about during this task. As you can see, most of their talking is devoted to making inferences, generating predictions, and commenting on connections to prior information. The relative frequencies of these activities are roughly the same across the four stories shown in table 5, though there are some interesting exceptions that we will comment on later. Though comments about the story or about their own understanding are low in overall frequency, they are very diagnostic of aspects of stories that readers are sensitive to.

We will first consider an example of a well-formed simple story in order to show how the principles bear upon what a reader is doing during comprehension. *Lentil* is a straightforward children's story whose surface organization is quite simple. The first 17 sentences are the Exposition, in which the three major characters and the seeds of the potential conflict are introduced. The Narration begins with sentence 18, and the climax that distinguishes Complication from Resolution occurs during sentences 30 to 32. Our analysis of what readers are doing while reading "Lentil" is based on twelve subjects who talked out loud while reading it. Later we will describe how the talking-out-loud data correspond to the sentence-by-sentence reading times of another group of subjects who were reading silently.

Throughout the Exposition of "Lentil" the talking-out-loud subjects were clearly collecting information and formulating tentative hypotheses about what was likely to happen in the story. They all recognized that the three central characters yield a

Table 5
Proportion of talking-out-loud protocols in each category.

Category	Lentil	Stranger	Ghosts	Circle Island
Predictions	0.22	0.23	0.13	0.26
Questions	0.04	0.02	0.10	0.01
Comments on structure	0.09	0.06	0.08	0.10
Comments on own behavior	0.03	0.03	0.04	0.02
Confirmation of predictions	0.02	0.03	0.01	0.02
References to antecedent information	0.29	0.36	0.30	0.27
Inferences	0.30	0.24	0.30	0.27
General knowledge and associations	0.02	0.04	0.03	0.06

highly probable line of conflict and resolution: Old Sneep and Colonel Carter will be the source of the conflict, and Lentil will provide the resolution. At sentences 18 and 19, where the Narration begins, the subjects brought together their tentative hypotheses and formulated general plans for the rest of the story. Table 6 shows some examples of what subjects said at this point. It was striking to us how regular this phenomenon was: virtually all subjects did the same thing at about the same place in the story.

Throughout the processing of the Narration the hypotheses constructed earlier by subjects were in evidence. Each event was incorporated into the general plan constructed at the end of the Exposition. The impact of these hypotheses was most dramatically revealed at sentences 30 and 31. The readers knew something was going to happen, but had no idea what. Consistent with the expectation that a story contains novel elements, the readers expected to be surprised by the specific

Table 6
Examples of reader comments to sentence 18 in "Lentil".

Subject	Comments
24	"I expect the plot will succeed in getting the Colonel to hear Lentil playing his harmonica. The Colonel will be impressed and Lentil will be rewarded somehow".
17	"Maybe a celebration is planned (parade, etc.) and Lentil will win the day with a rousing welcoming song".
07	"I expect to read that Colonel Carter now will have some kind of role in what's going on with Attu and his music and Old Sneep. I expect to hear something about Colonel Carter's reaction to Lentil".
23	"We expect to see some interaction between Colonel Carter, Old Sneep, and Lentil. Probably a great celebration. Suspect that Lentil will probably be asked to play for him".

form of the complication, even though they knew it would involve something that Old Sneep does to Colonel Carter. Sentence 30, "Then there was a wet sound from above" is totally surprising. No one can tell at this point what the wet sound is. But most subjects stated immediately, prior to going on to sentence 31, that Old Sneep must be implicated and that the climax is at hand. Even though Lentil had not been mentioned since sentence 14, most subjects predicted that Lentil's central role in the story was about to emerge. Thus, the "wet sound" is a highly informative event. Because the story is well-written (at least according to the conventions for such children's stories), the hypothesis-based predictions subjects made in response to the climactic event were in fact right. Once the climax is passed, the mode of processing changed quite drastically. The readers appeared to be operating in a confirmatory mode. No new surprises were expected. Over and over one finds in the protocols statements like "That's what I expected" or "Yup", indicating that predictions generated were now being found to be congruent with the emerging details of the Resolution.

"Lentil" is a good example of a story where the writer and the reader are operating in harmony. The Exposition leads all of the readers to establish a background and a set of hypotheses which successfully guide the subsequent processing. The same expectations are derived by different readers at just about the same points in the story, and specific events tend to be interpreted in the same way. The climax is especially striking. Though it is unpredictable, it is immediately interpretable and is integrated into the general representation of the story being constructed by the reader.

We have contrasted "Lentil" with several other stories, some of which clearly violate the principles listed in table 2. We will briefly mention two of these. The version of "Circle Island" used by Thorndyke (1977) has two serious violations of the principles. First, the whole story violated the principle in (2) in table 2, which, for stories, claims that the Narration of the story must be told in specific, concrete terms. In "Circle Island" no specific characters are developed and no specific actions or events are described. Most subjects in the talking-out-loud task assumed the text was the Exposition of a story. Many were surprised when it ended, stating that they expected a Narration to follow. To these readers the story was told at the wrong level. The second problem was that most subjects could not link the last sentence with the previous seventeen. The last sentence states that civil war broke out, but subjects consistently stated that not enough information had been presented to allow them to construct a reasonable scenario which led to a civil war. Thus, the principle of *Connectivity* was violated. Since "Circle Island" conforms to the story grammar described by Thorndyke, this is a good example of how adherence to a story grammar is not a sufficient condition for a story to be well-formed.

Bartlett's (1932) famous "War of the Ghosts" is not even well-formed at the level of underlying structure, at least for readers from our culture. Readers have a terrible time with it. There is no evidence that any of the twelve subjects who talked out loud in this story ever formulated a coherent hypothesis about the

global organization of the story, although some were able to generate some local hypotheses. The confusion of our subjects is illustrated in the comments in table 7. These comments were all made in reaction to the sentence in which ghosts are first introduced. In this sentence, one of the main characters infers that he has encountered ghosts. The basis for this inference is not at all clear to readers from our culture. The confusions generated by this sentence include low-level confusion about the referents of pronouns as well as confusion about the basis for the ghost inference itself.

Mandler and Johnson (1977) presented a convincing analysis which shows that the primary problem with "War of the Ghosts" is that readers have difficulty constructing any causal links between the individual episodes which follow each other in time. Thus, readers cannot discern an *Overall plan*, since neither *Coherence* nor *Completeness* can be detected. Further, there are too many bizarre or unfamiliar events, so readers have difficulty assimilating the story to a *Conventional world*. One can continue through the list of principles in table 2 and show that "War of the Ghosts" provides examples of violations of most of them.

By contrasting the talking-out-loud protocols of subjects reading well-formed and ill-formed stories, we have gathered considerable support for the claim that readers possess the kind of knowledge described in table 2. Further, these same data seem to reveal how subjects use this knowledge in their interpretation of each sentence as they encounter it in a story. However, given the obvious artificiality of the talking-out-loud task, is there any connection between what readers tell us in this task and what readers do in a task that more closely approximates normal reading? To answer this, we collected reading times for subjects who read each story silently

Table 7
Examples of reader comments to ghost sentence in "War of the Ghosts".

Subject	Comments
8	"I don't get that at all. I don't think it refers to the rest of the story".
9	"Is he the Indian? Somebody, either judiciously or seriously is suggesting that some of these could be ghosts. Could be the warriors, the young men, the violence".
12	"I can't imagine why he would say that, unless maybe they were speaking about him".
13	"(chuckle) Ah, well, the young man ... uh, I draw a blank on this sentence".
14	"Jees, I don't know what the hell is going on. They are afraid of somebody's ghost? We don't know who. One of the town people? Maybe the warriors are ghosts".
15	"This doesn't seem to fit in anywhere. I have no idea how to interpret this sentence or even where this sentence came from".
16	"The young man thought that the ... not sure what it means, but there were some ghostlike qualities to the Indian. I don't know what that means".

at a computer terminal. One overall effect in the reading times is that *War of the Ghosts* was read much more slowly than any of the other three stories. The average time per syllable was 415 msec., compared to 219, 278, and 180 for *Lentil*, *Stranger*, and *Circle Island*, respectively. This difference was highly significant in an analysis of variance ($F(3, 39) = 13.2$, $p < 0.001$). This fits in nicely with the overall general confusion shown by the talking-out-loud subjects who were reading *War of the Ghosts*.

We also examined the sentence-by-sentence details of the reading times. Fig. 3 shows the type of data provided by this reading time task. This figure presents the reading rates for a group of twelve subjects reading "Lentil". The reading rates shown on the ordinate of fig. 3 have been adjusted to take into account differences in sentence length. A qualitative examination of these reading times reveals some interesting correspondences with the talking-out-loud data. Notice that there are numerous long reading rates during the early portion of the story. These are most likely due to the attempts by subjects to store background information and formulate general hypotheses. Of particular note are the long times for sentences 16 through 20, where the Exposition shifts to the Narration. This is where the talking-out-loud subjects were formulating the general hypotheses they used to guide their further processing. The peak at sentence 31 and the flat reading times afterward correspond, respectively, to the large amount of inferential and predictive activity at the climax followed by confirmatory processing that we noticed in the talking-out-loud data. In general, the pattern of reading times corresponds to the reader's pattern of processing strategies revealed by our analysis of the talking-out-loud protocols.

These qualitative impressions for "Lentil" were supported by a quantitative analysis of the reading times. We performed a stepwise multiple regression using the

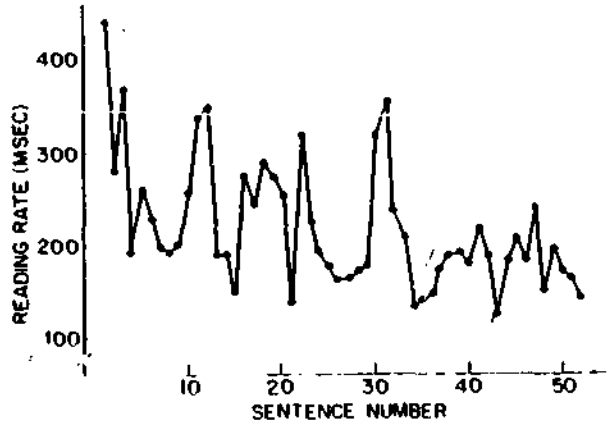


Fig. 3. Average reading rate for each sentence in *Lentil*.

mean reading time per sentence as the dependent variable and a series of possible predictors of reading time as the independent variables [1]. Table 8 shows the results of this analysis. This table shows the variables that were and were not selected by the stepwise multiple regression as significant predictors of reading time. Two analyses are shown: one that did and one that did not include the serial position of the sentence as a variable. Note that in both analyses the number of syllables per sentence accounts for a very large percentage of the variance in reading times. However, in addition, the number of inferences made by talking-out-loud subjects also emerged as a significant predictor, and in the second analysis where serial position was excluded, so did the number of predictions. What these analyses reveal is that there is an association between places in "Lentil" where the talking-out-loud subjects generate inferences and predictions and where the silent-reading subjects take more time. To us, this suggests quite strongly that the information we collect from the talking-out-loud subjects is useful in helping us to formulate hypotheses as to what the normal reader is doing while reading a simple story.

Tables 9 and 10 show similar multiple regression data for other stories we have examined. In table 9, another well-formed story shows similar associations between properties of the talking-out-loud data for that story and the corresponding reading times. In table 10, similar analyses for two ill-formed stories showed little association between the talking-out-loud data and the reading times. This difference between well-formed and ill-formed stories is potentially interesting, though we would want to replicate it for a wider range of stories before we would make much of it. But certainly the analyses of "Lentil" and "Stranger" are encouraging.

[1] Multiple regression analysis is a standard statistical technique for quantitatively evaluating the relationship between one or more predictor variables and a dependent variable. For instance, in tables 8, 9, and 10, reading time is the dependent variable, and a number of other variables are explored as possible predictors of the reading times across sentences. The terms "Coeff.", "Sig.", and "Cum. R^2" are statistics which pertain to the outcome and interpretation of this analysis. "Coeff." refers to the regression coefficients which indicates how many units of change occur in the dependent variable for each unit change in a given predictor variable (*e.g.*, in table 8, part 1, each additional syllable in a sentence is estimated to increase reading time by 194 msec). The "Sig." statistics indicate whether or not the effect of a predictor (indicated by the coefficient) is statistically significant (*i.e.*, is large enough not to be the effect of normal random variation). Selected predictors in these tables are defined as just those variables whose effects are statistically significant in this sense. Finally, the "Cum. R^2" statistics indicate what proportion of the total variation in reading time across sentences can be attributed to each predictor variable. For example, the variation in the number of syllables across sentences accounts for about 70% of the total variability in the reading times for these sentences, while serial position only accounts for an additional 5% of the variability. In general, the R^2 statistic is one measure of the relative importance of a predictor in the set, and the overall "Cum. R^2" provides an overall measure of how much variation in the reading times can be attributed to the set of predictors as a whole. Thus, in part 1 of table 8, about 79% of the total variability in reading times across sentences can be accounted for by the selected variables. The remainder is unexplained variation.

Table 8
Multiple regression analysis of reading times for "Lentil".

Forward stepwise regressions [a]	Coeff.	Sig.	Cum. R^2
1. *Predictors selected*			
Syllables	19.17	0.0000	0.696
Serial position	47.2	0.0019	0.748
Inferences	118.5	0.0036	0.789
Predictors not selected			
Importance			
Predictions			
Questions			
Comments on own behavior			
Comments on structure			
Confirmation of predictions			
Reference to antecedent information			
General knowledge and associations			
2. *Predictors selected*			
Syllables	189.7	0.0000	0.696
Inferences	122.8	0.0044	0.742
Predictions	87.9	0.0477	0.762
Predictors not selected			
Importance			
Questions			
Comments on own behavior			
Comments on structure			
Confirmations of predictions			
Reference to antecedent information			
General knowledge and associations			

[a] Dependent variable = mean reading time per sentence.

Processing simple essays

Because they are written and read for different purposes, one would expect stories and essays to be read in somewhat different ways. On the other hand, since both employ many general linguistic conventions, there might also to be many similarities. At a general level the similarities are captured by the kinds of principles shown in table 2. In this section, we want to focus on some of the differences we observed in how our subjects approached the reading of simple essays. To us, these differences clearly reveal that college readers know and use a wide variety of genre conventions in reading simple coherent texts.

We will first examine the data from our talking-out-loud task. Table 11 shows some general statistics about the frequency with which subjects said different types

Table 9
Multiple regression analysis of reading times for "Stranger".

Forward stepwise regressions [a]	Coeff.	Sig.	Cum. R^2
1. *Predictors selected*			
Syllables	212.9	0.0000	0.493
Serial position	59.5	0.0006	0.588
General knowledge and associations	223.9	0.0170	0.659
Predictors not selected			
Importance			
Predictions			
Questions			
Comments on structure			
Confirmation of predictions			
Reference to antecedent information			
Inferences			
2. *Predictors selected*			
Syllables	183.2	0.0000	0.493
Importance	23.3	0.0043	0.540
Predictions	89.5	0.0036	0.603
General knowledge and associations	211.7	0.0108	0.666
Predictors not selected			
Questions			
Comments on structure			
Reference to antecedent information			
Inferences			

[a] Dependent variable = mean reading time per sentence.

of things during the talking-out-loud task. The somewhat different distribution of statements in table 11 when compared with table 5 reflects the somewhat different instructions given to the talking-out-loud subjects in the essay tasks. Not surprisingly, explicitly asking subjects to talk about how the current sentence fit in to the overall plan of the essay generated more statements about the structure.

What is not revealed by these overall figures are some very important qualitative differences in what subjects said to essays when compared with stories. Perhaps the most dramatic difference came in the kinds of predictions made by readers. Note that in tables 5 and 11 predictions are relatively common: they represent about a quarter of the productions for the essays and a third for the stories. However, the kinds of predictions were very different. The typical predictions made by the reader of a story were specific. They predicted specific events that might occur later in the story, involving specific characters. Table 6 contains typical examples. The specificity of the predictions increased as the story developed, not sur-

Table 10
Multiple regression analysis of reading times for "War of the Ghosts" and "Circle Island".

Forward stepwise regressions [a]	Coeff	Sig.	Cum. R^2
"Ghosts"			
Predictors selected			
Syllables	358.8	0.0000	0.743
General knowledge and associations	513.9	0.0021	0.826
Predictors not selected			
Importance			
Predictions			
Serial position			
Comments on structure			
Reference to antecedent information			
Inferences			
"Circle Island"			
Predictors selected			
Syllables	182.4	0.0001	0.604
Predictors not selected			
Serial position			
Importance			
Predictions			
Questions			
Comments on structure			
Reference to antecedent information			
Inferences			
General knowledge and associations			

[a] Dependent variable = mean reading time per sentence.

Table 11
Proportion of talking-out-loud productions in each category for essays.

Category	Ice Ages		Carpet	
	Well-formed	Ill-formed	Well-formed	Ill-formed
Predictions	0.24	0.29	0.18	0.24
Questions	0.12	0.12	0.06	0.08
Comments on structure	0.32	0.24	0.28	0.32
Comments on own behavior	0.03	0.04	0.03	0.03
Confirmation of predictions	0.01	0.02	0.01	0.03
References to antecedent information	0.18	0.14	0.28	0.19
Inferences	0.10	0.12	0.11	0.10
General knowledge and associations	0.01	0.04	0.04	0.01

prisingly, since more and more information was available to the reader about the plot line being developed. But the important point was that the predictions were specific even very early on: those shown in table 6 were generated only about a third of the way into "Lentil". In marked contrast, the predictions given in the essays were *much* more general. They often consisted of comments to the effect that readers expected to see another argument, expected another example, *etc.* But the specific argument or the specific example were seldom generated. We think this difference in the types of predictions generated by readers reflects some important differences in the way they approach these two text types. The reader of a story has a set of firm expectations about the type of substantive events that will constitute acceptable continuations of a story. Given background information about the characters and their motives, much can be anticipated about how the conflict and resolution of the story will proceed. However, with an argument, it is apparently more difficult to predict the specifics. Readers know that an inductive argument (e.g., *Carpeting*) will have various elements of supporting evidence for the main point, and they know (perhaps trivially) that a compare and contrast will have comparisons and contrasts. But their predictions do not reduce to the level of specifics. For example, table 12 presents predictions made by subjects at the end of the first paragraph of the well-formed version of the *Carpeting* essay. The first paragraph gives some background information, introduces the thesis of the essay and announces that there will be three arguments made in support of the thesis. In contrast to the comments given at the end of the background section for *Lentil* (see table 6), the essay readers do not seem to be going beyond what the author has told them about the arguments to be presented.

Readers of the essays seemed to be quite sensitive to the surface devices used to organize the components of the text. This was revealed in two ways. First, readers seemed to be able to anticipate the ends of paragraphs and other major breaks in the argument quite well. Apparently they were sensitive to the level of the language and the general pace of the argument. Paragraphs were often organized by having a

Table 12
Examples of reader comments at the end of the introductory paragraph of the well-formed *Carpet* essay.

Subject	Comments
02	"It's going to go on to describe what is good about carpeting. Before it just told how it came to be that they're using more carpeting. I think they're going to tell the reasons why".
03	"I expect to find out what the three central arguments in favor of carpeting are".
04	"Now you would expect him to go on and possibly list and possibly explain each of the arguments in favor of carpeting".
06	"Sounds like they are going to sell me carpeting now. They will go into all the good points of carpeting and try to sell me a roll".

general claim or point made near the beginning (*i.e.*, a topic statement) and then following this with a variety of specific points or illustrations. Often the paragraph would end with a general summary. This alternation of general and specific language is clearly a convention of essay writing. Readers expect general statements of main points. They don't expect the main points to be implicit. However, they expect the main points to be supported with concrete examples or arguments. Further, they expect only so much supporting detail. Though an inductive argument and each major supporting point potentially have an indefinite amount of material, writers tend to present only a small number of what is possible: two or three major points, one or two examples, *etc.* Readers expect this, and use these expectations to pace their understanding of the text. These general conventions are described in rhetoric books (*e.g.*, Baker 1976; Hughes and Duhamel 1962; Kinneavy 1971; Payne 1969), and our readers were sensitive to them.

The readers were also sensitive to surface rhetorical devices used to organize the text. In the reading time profile for the well-formed version of *Carpeting* (shown in fig. 4), two of the longest reading times are for sentences beginning with the rhetorical device, "in short" (sentences 15 and 29 in fig. 4). Recall that subjects were told to read with the goal of writing a summary later. The reading time pattern suggests to us that subjects were using rhetorical devices to signal where they should focus their attention in order to be able to write a good summary later on. No such signals were available in the ill-formed version of the *Carpeting* essay. Performance in the talking-out-loud task suggests that this lack of signalling had an effect. While readers of the well-formed version always had a good sense of where they were in the essay, there was much evidence in the ill-formed version that readers were often confused or misled as to the progress of the argument.

The readers of the essay on the ice ages showed their sensitivity to the organiza-

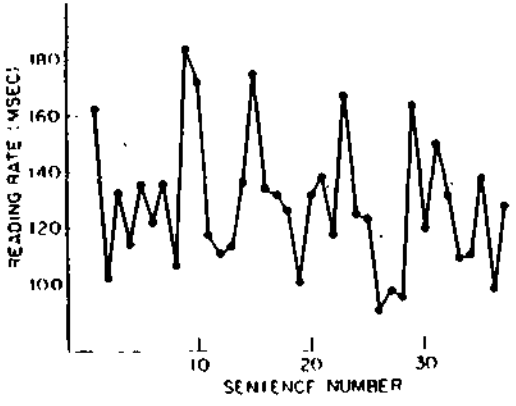


Fig. 4. Average reading rate for each sentence in the well-formed version of *Carpeting*.

tion in another way. The two versions of this essay varied on where the general conclusion was stated. Those who had the conclusion stated only at the end had a much harder time getting the drift of the essay. Their protocols revealed consistent confusion as to the central point of what they were reading. Again, readers expect to be told what the essay is about early on, and are confused when this does not happen.

The reading times were sensitive to the well-formedness of the essays. For both versions, the overall reading times were considerably slower for the ill-formed versions than for the well-formed ones. Specifically, a regression analysis that partialled out variables like sentence length or syllables revealed no difference in overall reading time per sentence for *Ice Age* or *Carpeting* essays, but the ill-formed version of each essay required about 900 msec longer to read per sentence than the well-formed versions ($F(1, 113) = 7.05, p < 0.001$). Well-formed versions of *Ice Age* and *Carpeting* essays took 3680 and 3773 msec per sentence, respectively, while the ill-formed versions took 4576 and 4387 msec per sentence.

Another difference between the essay data and the story data was in the relationship between the talking-out-loud protocols and the reading times. A series of multiple regression analyses revealed no clear relationship between any quantitative characteristic of the talking-out-loud data and the mean reading times. The only variable to emerge as a significant predictor of reading time was the length of the sentence. This is in marked contrast to the story data, where several indices of talking-out-loud behavior were correlated with the reading times. This is yet another indication of the differences in the behavior of the subjects who were reading the two genres. While the subjects talking-out-loud to the stories were generating rich, interconnected hypotheses as they progressed through the passage, those reading the essays did little more than comment on general aspects of the essay structure. This showed up in the emergence of serial position of a sentence as a factor in the multiple regressions for the stories but not for the essays. Seemingly, subjects read more slowly at the beginning of a story while they generated hypotheses and more quickly as they got farther into the story because much of what they were doing was confirming earlier predictions. No similar pattern emerged in the reading times for the essays, nor was there evidence in the talking-out-loud protocols of a similar predictive strategy. Story readers' hypothesis construction seemed to lead to different amounts of time being devoted to different story constituents, as revealed by the intercorrelations of the talking-out-loud and reading time tasks. No similar process seemed to be involved in the reading of the essays.

To summarize, the reader of an essay has general expectations about the overall structure of the argument, comparison, or other point being made in the essay. The reader quickly recognizes the type of point being made, and at a general level is sensitive to the organizing devices in the surface structure of the essay. However, unlike the story reader, the reader of an essay does not appear to engage in rich hypothesis generation and testing. The reader seems to adopt a more passive strategy, waiting for each new item of information to be presented and trying to fit it into the overall scheme of the argument.

Of course, one ought to be cautious in overinterpreting the behavior of readers from the small sample of texts we have examined. In this paper we have reported on four stories and two essays (each with two versions), and any literate person knows that the domains of stories and essays are very rich, with many types. We are actively pursuing this issue by examining a much wider range of texts in both domains. But we feel relatively confident that the broad differences between story readers and essay readers that we have described in this paper are accurate characteristics of literate readers.

General discussion

In this paper we have reported on some results obtained for two simple genres, stories and essays. These are genres that are familiar to any reasonably literate reader, such as the college students who served as our subjects. Our main claim about these readers is that they know, either explicitly or tacitly, a number of conventions for how texts of these types are written, and they use this knowledge actively during their comprehension of the text. They have strong expectations about what they will find in the text, and use these expectations to guide their understanding. When their expectations are violated, as in an ill-formed story like "War of the Ghosts" or the ill-formed versions of our two essays, their verbal protocols reveal confusion and their silent reading times are slowed down considerably. When their expectations are met, as in "Lentil", "Stranger", or the well-formed versions of our essays, they appear to be able to profit from the application of their knowledge of genre conventions to the process of understanding.

A capsule summary of the difference in strategies for the readers of stories and essays might go as follows. The basic orientation of the reader to a story is *prospective*. The reader is looking ahead, trying to anticipate where the story is going. Except at the beginning, where an overall hypothesis is being developed, the story reader tends to relate each sentence to the general hypotheses and predictions that have been developed. In contrast to this, the reader of the essay appears to adopt a *retrospective* orientation. Each new element in the essay is related to earlier elements. There is little anticipation of what is coming up, except at the most general level. This difference in orientation on the part of the reader is of course due to the basic difference in underlying structure of these two genres. The story has a causal, temporal structure, with events ordered and interrelated in well-specified ways. Further, the general schema for a story dictates that the events must unfold according to a pattern of complication and resolution. These constraints make a predictive strategy quite useful. An essay, such as an inductive argument, has no similar structure. Rather, a general hypothesis is supported by a variety of evidence, and the support relationship between each succeeding piece of evidence and the (usually) previously stated thesis makes the retrospective strategy appropriate. It is extremely interesting to us that our readers, especially those in the talking-out-loud task, exploited these differences in establishing their overall strategy for comprehending the texts.

If our talking-out-loud data are representative, readers appear to have considerable knowledge about both the underlying form and surface conventions for these simple genres. Table 2 attempted to capture at a general level the kind of knowledge they have. Presumably this general knowledge represents a kind of genre schema, and would be applicable to a wide variety of specific text types. While the schema sketched in table 2 has obvious limits (e.g., it has limited utility for poetry), it probably reflects the general expectations literate people have for many types of texts. We believe that our empirical investigations provide a good start at discovering what people know about two simple types of discourse. However, we are obviously far from a complete cognitive theory of genre. In the remainder of this discussion we would like to raise four issues which need to be addressed if we are to develop a more complete theory of genre.

First, we need to characterize in much more detail the principles in table 2. For example, what is an adequate formal or linguistic representation of these principles? Can we develop a plausible psychological representation for the principles? Do the principles all have the same status? We have already suggested that the principles of *Connectivity* and *Overall plan* could be formally represented in terms of, say, a story grammar representation (although the specifics and adequacy of such representations is a matter of considerable debate). Other principles like *Specificity* or *Language* do not seem to have the same status, but refer rather to more qualitative stylistic or aesthetic considerations. Specifying the formal and psychological nature of these principles is especially important if we want to characterize precisely the well-formedness of texts.

A second issue is how we can apply or instantiate the general principles given a particular genre. At present the principles have been applied on intuitive grounds, using empirical protocols as a heuristic to illustrate or suggest a particular formulation of a principle for a particular genre. However, we obviously need a more principled way of applying the principles.

A third issue is how these principles are learned, not only in children learning to read and write particular genres but also in adults who may have occasion to master new forms of discourse in a profession or elsewhere (e.g., technical forms of writing, instructions, memos, new literary genre, and so on). Presumably, the adult reader would bring to bear something like the principles in table 2 and would induce from particular instances of a genre how these principles are particularized.

Finally, while our analysis of written communication examined the roles of both the reader and writer, we have focussed in this paper on the reader's behavior. But our analysis – in particular, the principles listed in table 2 – is just as applicable to the writer. Presumably a writer must learn both the underlying forms for a genre and the principles of surface composition. In fact, in the abstract the contrast between underlying and surface forms could correspond to the generation and the revision processes often discussed in connection with writing (e.g., Grey 1972;

Flower and Hayes 1977, 1979). Good and bad writers ought to differ, in part at least, in the extent to which they have acquired the principles of composition for the genres they plan to write. However, while they may have tacit understanding of the principles in their role as a reader, these do not necessarily readily translate into effective writing strategies. A significant part of the writer's task is to be able to anticipate the effects of his or her composition on the reader. Failure to do so effectively usually results in a poor composition. Studies of the contrast between good and bad writers, of writers who are studying composition, and of an individual writer progressing through successive drafts ought to be revealing of the extent to which these principles play a role in the writer's task.

We have found our talking-out-loud task to be quite revealing of both the knowledge possessed by readers and the processes they employ to understand a simple text. In this paper we have only conveyed a very little of the information contained in the protocols for this task. We believe this technique has much potential usefulness for gaining and understanding of the conventions governing a wide range of text types. Presumably the task could be of great interest to scholars of literary processes as well as those, like us, who are primarily interested in the cognitive processes of ordinary readers.

References

Baker, S. 1976. The complete stylist and handbook. New York: Thomas Y. Crowell.

Bartlett, F.C. 1932. Remembering: an experimental and social study. Cambridge: Cambridge University Press.

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

Bransford, J.D. and M.K. Johnson. 1973. 'Considerations of some problems of comprehension'. In: W.G. Chase, ed., Visual information processing. New York: Academic Press.

Brewer, W.F. and J.H. Lichtenstein. 1980. Event schemas, story schemas, and story grammars. Paper presented at Attention and Performance IX, Jesus College, Cambridge.

Cirilo, R.K. and D.J. Foss. 1980. Text structure and reading time for sentences. Journal of Verbal Learning and Verbal Behavior 19: 96–109.

Dawes, R. 1966. Memory and the distortion of meaningful written material. British Journal of Psychology 57: 77–86.

Dooling, D.J. and R.L. Christiaansen. 1977. 'Levels of encoding and retention of prose'. In: G.H. Bower, ed., The psychology of learning and motivation. Vol. 11. New York: Academic Press.

Finlay, W. 1969. Folk tales of the north. New York: Watts.

Flower, L.S. and J.R. Hayes. 1977. Problem-solving strategies and the writing process. College English 39: 449–461.

Flower, L.S. and J.R. Hayes. 1979. A process model of composition. Technical Report No. 1, Document Design Project, Carnegie-Mellon University.

Gary, B.J. 1972. The writing process. Belmont, CA: Wadsworth.

Hirsch, E.D., Jr. 1977. The philosophy of composition. Chicago, IL: University of Chicago Press.

Hughes, M. and P.A. Duhamel. 1962. Rhetoric: principles and usage. Englewood Cliffs, NJ: Prentice-Hall.

Johnson, N.S. and J.M. Mandler. 1980. A tale of two structures: underlying and surface forms in stories. Poetics 9: 51–86.

Jones, A.P. and C.W. Faulkner. 1968. Writing good prose: a structural approach to writing paragraphs and themes. 2nd. ed. New York: Scribner's.

Kieras, D.E. 1978. Good and bad structure in simple paragraphs: effects on apparent theme, reading time, and recall. Journal of Verbal Learning and Verbal Behavior 17: 13–28.

Kinneavy, J.L. 1971. A theory of discourse. Englewood Cliffs, NJ: Prentice-Hall.

Kintsch, W. and T.A. van Dijk. 1978. Toward a model of text comprehension and production. Psychological Review 85: 363–394.

Klotsch, W. and D. Vipond. 1978. 'Reading comprehension and readability in educational practice and psychological theory'. In: L.W. Nilsson, ed., Memory: processes and problems. Hillsdale, NJ: Lawrence Erlbaum.

Mandler, J.M. and N.S. Johnson. 1977. Remembrance of things parsed: story structure and recall. Cognitive Psychology 9: 111–151.

McCloskey, R. 1940. Lentil. New York: Scholastic Book Services.

Meyer, B. 1975. The organization of prose and its effect on memory. Amsterdam: North-Holland.

Miller, J.R. and W. Kintsch. 1980. Readability and recall of short prose passages: a theoretical analysis. Journal of Experimental Psychology: Human Learning and Memory 6: 335–354.

Minsky, M. 1975. 'A framework for representing knowledge'. In: P. Winston, ed., The psychology of computer vision. New York: McGraw-Hill.

Olson, G.M. In preparation. The concept of a story.

Olson, G.M., S.A. Duffy and R.L. Mack. 1980. 'Applying knowledge of writing conventions to prose comprehension and composition'. In: W.E. McKeachie, ed., Learning, cognition, and college teaching. San Francisco, CA: Jossey-Bass.

Olson, G.M., R.L. Mack and S.A. Duffy. In preparation. Composition and comprehension.

Payne, L.V. 1969. The lively art of writing. New York: New American Library.

Piaget, J. 1952. The origins of intelligence in children. New York: Norton.

Pratt, M.L. 1977. Toward a speech act theory of literary discourse. Bloomington, IN: Indiana University Press.

Ricoeur, P. 1976. Interpretation theory: discourse and the surplus of meaning. Fort Worth, TX: Texas Christian University Press.

Rubin, A.D. 1980. 'A theoretical taxonomy of the differences between oral and written language'. In: R.J. Spiro, B.C. Bruce and W.F. Brewer, eds., Theoretical issues in reading comprehension. Hillsdale, NJ: Lawrence Erlbaum.

Rumelhart, D.E. 1975. 'Notes on a schema for stories'. In: D.G. Bobrow and A. Collins, eds., Representation and understanding: studies in cognitive science. New York: Academic Press.

Rumelhart, D.E. 1978. 'Understanding and summarizing brief stories'. In: D. LaBerge and J. Samuels, eds., Basic processes in reading: perception and comprehension. Hillsdale, NJ: Lawrence Erlbaum.

Rumelhart, D.E. and A. Ortony. 1977. 'The representation of knowledge in memory'. In: R.C. Anderson, R.J. Spiro and W.E. Montague, eds., Schooling and the acquisition of knowledge. Hillsdale, NJ: Lawrence Erlbaum.

Schallert, D.L., C.M. Kleiman and A.D. Rubin. 1977. Analyses of differences between written and oral language. Technical Report No. 29. Urbana, IL: University of Illinois, Center for the Study of Reading.

Schank, R.C. 1979. Reminding and memory organization: an introduction to MOPs. Research Report No. 170, Department of Computer Science, Yale University.

Schank, R., and R. Abelson. 1977. Scripts, plans, goals and understanding. Hillsdale, NJ: Lawrence Erlbaum.

Scholes, R. and R. Kellogg. 1966. The nature of narrative. New York: Oxford University Press.

Shaughnessy, M.P. 1977. Errors and expectations. New York: Oxford University Press.

Stein, N.L. and C.G. Glenn. 1980. 'An analysis of story comprehension in elementary school children'. In: R. Freedle, ed., Advances in discourse processes: discourse processing: new directions. Vol. 2. Norwood, NJ: Ablex.

Sternberg, M. 1978. Expositional modes and temporal ordering in fiction. Baltimore, MD: Johns Hopkins University Press.

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

Thornley, W.R. 1976. Short story writing. New York: Bantam.

Wilensky, R. 1978. Understanding goal-based stories. Research Report No. 140, Department of Computer Science, Yale University.

Wilensky, R. 1980. Points: a theory of story content. Memorandum No. UCB/ERL M80/17, Electronics Research Laboratory, University of California, Berkeley.

Winograd, T. 1972. Understanding natural language. New York: Academic Press.

Gary M. Olson is an Associate Professor of Psychology at the University of Michigan. A Stanford Ph.D., his main interests are in memory, language, and complex cognition.

Robert L. Mack is a Postdoctoral Fellow at I.B.M.'s John Watson Research Center in Yorktown Heights, New York. He recently received his Ph.D. from the University of Michigan, and is interested in language comprehension processes.

Susan A. Duffy is a doctoral candidate at the University of Michigan. She has a B.S. from Radcliffe and an M.Ed. from Harvard, and is interested in language comprehension processes.

Thinking-out-loud as a Method for Studying Real-time Comprehension Processes

Gary M. Olson,
University of Michigan
Susan A. Duffy
University of Massachusetts
and
Robert L. Mack
IBM Research Center

The analysis of cognitive processes in real time is one of the most methodologically difficult tasks in all of psychology. The events we wish to examine are internal to the mind, with only occasional observable correlates. Further, most cognitive tasks involve a host of hierarchically interrelated subcomponents, likely operating in parallel. Reading text in order to understand it is an excellent example of just such a task. And yet, there is increasing recognition of the fact that a deep understanding of how to assess the readability of texts and how to remedy reading difficulties will require an analysis of the process of comprehension (e.g., Kintsch & van Dijk, 1977; Olson, Mack, & Duffy, 1981).

Though many psychological processes important to comprehension occur outside of awareness, any sophisticated reader is aware of much cognitive activity that occurs during reading. With this in mind, we felt that one simple strategy for obtaining information about the process of comprehension would be to have readers think out loud while reading. We were motivated by a belief that intelligent reading has many affinities with problem-solving, a domain in which thinking-out-loud (TOL) protocols have proved to be a useful research tool (e.g., Newell & Simon, 1972). Of course, to use TOL data to study cognitive processes, one must be aware of the limits and pitfalls of this method. As with any other method, it is useful for pursuing some goals and not others. The aim of this chapter is to discuss its usefulness as one technique for studying the comprehension of connected text.

This chapter is organized as follows. The general rationale for the use of the TOL method is described first. This includes a discussion of a general model of comprehension that has guided our research and our use of this method. The goals of the TOL method and the general assumptions made in using it are described. Next, we list a variety of different types of TOL tasks that can be used, and briefly discuss their virtues. Then we illustrate the use of TOL data in the analysis of comprehension processes by discussing a series of studies conducted in our laboratories. We also describe a series of other applications of the TOL task to make comprehension activities explicit, though in some cases as a means to another goal. Since other investigators have used the TOL task with more mixed results than ours, we discuss some of the reasons why the TOL task may vary in its usefulness. In the concluding section, we summarize in a convenient form the advantages and the limitations of using TOL protocols for studying language

comprehension.

I. THE USE OF THINKING-OUT-LOUD PROTOCOLS

What kind of information can we hope to obtain from TOL data? There has been much controversy in the history of psychology about verbal reports as data, focussing primarily upon their oft-reported unreliability (e.g., Nisbett & Wilson, 1977). In a detailed analysis of verbal reports as data, Ericsson and Simon (1980) clarified several points about their use that are important to keep in mind when thinking about TOL data. First, the focus of the TOL task should be to get subjects to report the content of their immediate awareness rather than to report explanations of their behavior. Further, subjects should be asked to report what they are thinking about right now, not what they remember thinking about some time ago. The TOL task should also have subjects talk about aspects of their immediate experience that they can talk about. Some processes are unavailable to introspection or are difficult to verbalize. In general, limits on what is available to be reported upon, what can be remembered, and on the human ability to offer explanations or justifications for one's own behavior should be respected.

Furthermore, TOL data should not be taken as direct reflections of thought processes but rather as data which are correlated with underlying thought processes. TOL data provide a sample of what's on the subject's mind during the task. But they will not necessarily reveal the strategies, knowledge sources, or representations actually used. These theoretical constructs must be inferred from the TOL data. The situation is quite analogous

to the use of eye movements or reading times in studying comprehension. We are less interested in the statistical properties of eye movements or reading times than we are in the comprehension processes which generate these properties. These processes must be inferred from the data. TOL protocols are no different. They are unlikely to reveal in a direct fashion the underlying processes we are most interested in discovering.

These cautions are extremely important, for as Ericsson and Simon (1980) point out in some detail, many of the criticisms of verbal reports as data are based on faulty assumptions about the reasonable use of such data. As with any other form of data we collect in cognitive research, TOL data provide indicators of real-time processes that must be affirmed through the examination of as broad a range of different measures as possible. Of course, as with any other type of data, TOL data have a number of limitations which must be keep in mind. Ericsson and Simon (1980) present a clear discussion of the virtues and limits of TOL data in general, and this chapter will attempt to do the same for the specific case of text comprehension.

Reading involves a broad array of processes, from sensory and perceptual ones to higher level processes such as reasoning and inference. Table 1 presents a partial list of these. The perceptual, attentional, and memory processes involved in the recognition of letters and words for a skilled reader occur too rapidly and may not be available to consciousness. There are a variety of experimental procedures available for the analysis of these lower level processes. Such processes as the syntactic and

semantic analysis of sentences may or may not be usefully analysed by TOL methods. But the TOL task is best used to study the higher level processes in reading: the inferences, predictions, schema elaborations, and other complex cognitions that occur as part of skilled reading. We assume these processes are most available to consciousness as the reader reads. The outputs of these processes are verbal, slow to arise, and samples of them are sufficient for the investigator to infer what must have transpired. In general, to the extent that one agrees with Neisser's (1967) characterization of reading as "externally guided thinking [p. 136]," the TOL method is specialized for the study of the thinking.

Insert Table 1 about here

The investigation of reading processes can begin in many ways. Many have started with a formal analysis of the materials themselves. Those interested in letter or word recognition attend to the frequency and regularity of various items in typical text. Those interested in syntactic analysis turn to linguistic theories of sentence structure for hypotheses. Finally, those interested in text level variables look to linguistic analysis for descriptions of inter-sentence phenomena like anaphora or of the overall structure of texts. The information gleaned from such formal analyses of the properties of print and text is very important to the analysis of reading, and in combination with assumptions about psychological processes, provides a rich source of initial hypotheses about aspects of the

reading process.

Formal analyses of the properties of print and text are much less useful in formulating hypotheses about the higher level processes--the thinking--that occurs as part of skilled reading. Further, although many of us are aware of the thoughts we have while reading, it is difficult to analyze these processes either on the basis of introspection or from other a priori considerations. The TOL task offers an opportunity to collect systematic observations about the thinking that occurs during reading, allowing the investigator to form hypotheses about this level of processing which can in turn be evaluated in a number of ways, including experimental tests. Thus, the optimal use of the TOL task is as a discovery procedure for studying these higher level processes.

In order to better appreciate what might be learned about the comprehension process from TOL data, we begin with a brief sketch of the nature of text comprehension. Figure 1 shows a scheme for what is involved in reading, focussing on the higher level activities. The focus is on what the reader is doing at some particular point in the text. What controls the reader's thought processes at the point where sentence N in a text is being read? It is controlled by the structure of the text and sentence N's particular role in this structure. And it is controlled by the reader's knowledge of the world--both physical and social--that plays such an important part in language understanding in general. Further, it is quite clear that skilled readers possess knowledge about text conventions, about

how texts are written in order to accomplish what they perceive to be the author's goals (see Olson, Duffy, & Mack, 1980; Olson, Mack, & Duffy, 1981).

Insert Figure 1 about here

To these general sources of constraints are added three types of knowledge constructed by the reader during comprehension. The first is a representation of what has been presented in the text so far. This is typically organized by a scheme appropriate for the text type being read. Second, a workspace containing current lines of thought the reader is working on is constructed. In stories, the workspace might contain hypotheses about where the text is heading. For argumentative essays or journal articles, the workspace might contain criticisms of the author's arguments. The contents of the workspace can be more or less specific depending on a variety of reader and text variables (e.g., Olson et al., 1981), but they are an important part of what most skilled readers are doing in any text. Finally, although it has not yet been studied in much detail by cognitive psychologists, readers probably construct a model of the writer, the other participant in the social intercourse being mediated by the text. The reader's model of the writer may not be accurate, and may well be manipulated intentionally by the writer.

A reader facing sentence N in a text will use these various sources of knowledge along with a semantic representation of the sentence to modify each of the three representations being

constructed during comprehension: the representation of what has been read, the representation of hypotheses, and the representation of the writer. For conceptual simplicity we imagine that a sentence that has already been given a semantic representation is passed on to those processes most directly concerned with this updating (We ignore the undoubtedly important interactions between these updating processes and the lexical, syntactic, and semantic analyses of individual sentences). The major focus of the TOL data will be on the processes responsible for integrating the semantic representation of an individual sentence into the various cognitive structures being constructed during comprehension. These data should reveal the kinds of strategies used by readers in accomplishing these tasks, the kinds of knowledge sources employed, and the kinds of representations constructed. While memory measures like recall have provided useful information about the knowledge sources and representations used in text comprehension, they tell us very little about the strategies employed or about the sentence-by-sentence interactions among the knowledge sources and representations.

2. TYPES OF TALKING-OUT-LOUD TASKS

There are many kinds of TOL data that can be collected for studying text comprehension. In this section we will review several major types that have actually been used, and briefly discuss their virtues and limits.

1. Sentence-by-sentence talking.

In this version of the task, which could be considered its

most basic form, the subject is asked to talk after each sentence in the text. The talking continues until the text is completed. Though in principle the text could be presented in almost any fashion, most investigators have presented it in such a way that the reader cannot look ahead. However, investigators have varied the extent to which the reader can look back at previous text. The most restrictive presentation makes only the current sentence available, while the least restrictive makes all the previous text available. Various windows of intermediate size could also be used. How exactly to organize the presentation of the text depends somewhat on the goals of the investigator. For instance, in much of our work we have been interested in explicating as fully as possible the role of the current sentence in comprehension. This has led us to use the single-sentence exposure most frequently, but other arrangements can be easily justified.

Another dimension of variation is what the readers are asked to talk about. We have used two types of instructions in our work:

a. General instructions. In this version of the task, we encourage subjects to talk about a wide range of things. However, we typically give them a list of examples of the types of things we would like them to talk about. We feel it is inadequate to instruct them to "think out loud" without telling them what this means, at least in the context of studying comprehension.

We have done several studies in which we asked subjects to

think out loud generally (e.g., Olson et al., 1981). Keep in mind that these subjects are talking after each sentence in the text. The kinds of things we asked them to talk about included any inferences or elaborations they felt compelled to draw on the basis of the current sentence, any connections they saw between the current sentence and any prior ones, any predictions they had about what might be coming up, and any comments they had about what they felt was the role of the current sentence in the overall organization of the text. These are obviously not the only things subjects could talk about. We included these because they were of theoretical interest to us. The important point vis-a-vis the use of the TOL task to study comprehension is that one be explicit with the subject about what to talk about. The exact list of suggestions should be motivated by theoretical ideas or by prior research.

b. Focussed instructions. In this version of the TOL task, the subject is asked to talk about only one type of thing or to do one type of activity. Let us illustrate this with two examples from our own research. Because of our interest in predictive processing, we have collected TOL data in which all we asked subjects to do was to make predictions about what was coming up in the future in this text and to comment on the fate of earlier predictions if they felt this was warranted. Clearly, this task was motivated by a special theoretical concern for the importance of a particular type of activity. Earlier data from a general TOL task had indicated that predictions might be an especially informative kind of activity (Olson et al., 1981), so

we wanted to collect a richer set of prediction data that we had obtained from the general task, where subjects talked about a number of different types of things which compared with predictions.

In a second example, we had subjects ask questions after reading each sentence. They were told to imagine that the text's author was present, and that the author was willing to answer any question the reader had about the text at that point, except for the obvious question of what comes next. Once again, we had a specific theoretical goal in using this task. As with the example of predictions, these kinds of questions occasionally appeared in the more general type of TOL task, but we wanted a richer set of data than we could get when a variety of different types of information were being collected.

Both of these examples share several important properties. First, the selection of the thing to be talked about was theoretically motivated. In both cases we had a background of general TOL data, reading time data, and theoretical models from which we arrived at the specific probes we used. Second, the main reason for using the focussed TOL task was to get richer data about this particular data type. TOL subjects will only say so much. When you have a variety of types of things for them to talk about, you will only get a modest amount of any one type. In order to get richer data of a specific type the focussed task is needed. But because it is focussed it provides a much narrower window into the process of comprehension. Later, when we discuss an example of this methodology in more detail, we will return to

this issue.

r. Other variants. Rumelhart (1981) and Greesser (1981) have both used a TOL task in which subjects answered specific questions after each sentence of a text. Rumelhart's subjects answered five Wh-questions (who, what, why, when, where) after each sentence. The texts began ambiguously, with each successive sentence providing additional constraints on what was going on. Rumelhart used the TOL task to discover how subjects developed an interpretation of the texts and how the interpretation changed with each additional sentence. Greesser has used a variant of this task (subjects actually wrote down their answers) to investigate how readers construct a representation of the text, focussing on the contribution each sentence makes to the growing representation.

2. Selective talking.

Yet another variation on the TOL task is to have subjects talk at only particular points in a text. There are two broad classes of justifications for this. First, one may have a process theory that pinpoints certain places in a text as crucial tests of some aspect of the theory. Second, in designing materials for an experiment, there may be certain properties one wants to have at certain points in the experimental texts. A selective TOL task can be used to verify that the materials have in fact instantiated this property.

Except for the fact that it is selective rather than sentence-by-sentence for the whole text, the earlier discussions about what to have subjects talk about hold for this variation of

the task too. We would especially like to underscore the use of this version of the TOL task for preparing stimulus meterials for experiments. In most studies of text processing, the major independent variables are manipulations of the text. We have found through experience that what seem to be instantiations of a text variable are often not good examples of it. Using selective TOL to aid in developing the stimuli has been a major part of several studies we have done. The feedback from subjects who think out loud is extremely informative, so much so in fact that we have begun a whole line of research aimed at using TOL data as tools for providing feedback to writers. We will discuss this last example later.

In general, selective talking is used to study what processing is like at some particular point. However, to assess the role of local contributions to the talking versus global characteristics of the talking, the placement of control probes at points in the text different from those of specific interest is important. A variety of theoretical considerations would determine where to put such control probes, but they are a necessary part of the use of selective talking.

3. After the fact talking.

Ericsson and Simon (1980) correctly stress that it is risky to ask subjects to talk about their cognitive experiences after the fact. Memory is too fallible to allow for accurate reporting of earlier mental states. However, if very short texts or text fragments were used, the memory problems are not as great, and useful TOL data could be collected after the fact.

Collins, Brown, and Larkin (1980) used just such a task to examine the general strategies used to comprehend short texts (3 and 4 sentences long). The texts were difficult to understand, and subjects were asked to talk about the hypotheses they had considered and rejected in trying to interpret the text. Because the texts were short, subjects could remember intermediate interpretations they had generated while reading. This method is less useful for exploring the processing of longer texts, especially if the investigator is concetned with evaluating the contribution each sentence makes in the comprehension process. Even with the short texts used by Collins et al., it is not clear how accurate subjects were in pinpointing exactly where a hypothesis was introduced or rejected.

Summary

Our descriptions of different TOL tasks in this section by no means exhaust all of the possibilities. Rather, these should be taken as suggestive of the kinds of ways in which TOL data can be collected to be used to explicate the nature of comprehension processing. One's specific theoretical and empirical goals will dictate which variant of the TOL task will be most useful.

3. EXAMPLES OF THINKING OUT LOUD DATA

In this section we present some examples of the use of TOL data from our own research. We present an example of both the general TOL sentence-by-sentence task and the focussed sentence-by-sentence question-asking task.

Sentence-by-sentence TOL for simple stories and essays.

These data, reported in more detail in Olson et al. (1981),

illustrate the use of the general sentence-by-sentence TOL task. We had subjects think out loud while reading four different stories and four different essays. We used two well-formed simple stories, Lentil and Stranger, and two stories that violated certain conventions of story telling. These latter two stories were Circle Island, Dawes' (1966) story that was used by Thorndyke (1977) to study story processing, and War of the Ghosts, Bartlett's (1932) classic. We chose Circle Island and War of the Ghosts because we wanted to be able to contrast TOL data obtained from well-formed stories (Lentil and Stranger) with these two ill-formed stories. A different group of subjects talked out loud to two versions of each of four different essays. We present the story data in more detail here, so will not describe the essays or their data as completely (see Olson et al., 1981, for details).

It helps to have a sense of what TOL data are like. Thus, in Table 2, we present an excerpt from a subject talking out loud to the first 17 sentences of Lentil (All of Lentil is shown in Table 3). These data are typical of what a skilled, educated adult reader talks about while reading a story of this type.

Insert Tables 2 and 3 about here

We have used protocol data of this type in two general ways. First, we have obtained qualitative impressions of the nature of comprehension processing for various types of texts. Second, we have related quantitative properties of the TOL data to other types of data, such as sentence-by-sentence reading times and

recall data. We give examples of each type of analysis here to illustrate how TOL data can be used.

Qualitative analyses. We focus on TOL data obtained for simple stories. What is a reader doing while reading such a text? The reader is confronted with the following task. A text embodying various aspects of the story is available as input. The text will be processed against a background of several types of knowledge that are relevant to it. The final product of understanding will be an interpreted representation of the essential elements of the story. However, with most stories, this representation must be constructed from incomplete information in the text. Much that is important to interpreting the elements of the text as a story is left implicit.

We can make this more concrete by describing some of the properties of TOL data for readers reading simple stories. Table 4 presents the relative frequencies of different types of things readers talked about during the four stories we described earlier. As is evident, most of their talking is devoted to making inferences, generating predictions, and commenting on connections to prior information. This follows from the stress we gave these three categories of information in our instructions to subjects. The relative frequencies of these activities are roughly the same across the four stories shown in Table 4, though there are some interesting exceptions that are discussed in Olson et al. (1981). Though comments about the story or about their own understanding are low in overall frequency, they are very diagnostic of aspects of stories that readers are sensitive to.

Insert Table 4 about here

A portrait of story processing is revealed by readers reading a well-formed story, Lentil (see Table 3). This is a straightforward children's story whose organization is quite simple. The first 17 sentences introduce the three major characters and lay the seeds of the later conflict among them. The detailed actions of the story begin at sentence 18, and the climax that distinguishes the complication in the plot from the resolution occurs during sentences 30 to 32. Our analysis of what readers are doing while reading Lentil is based on 12 subjects who talked out loud during it. We were especially concerned to identify aspects of their talking that were common across most of the group rather than idiosyncratic to individual subjects.

Throughout the first 17 sentences the subjects were clearly collecting information and formulating tentative hypotheses about what was likely to happen in the story. They all recognized that the three central characters yield a highly probable line of conflict and resolution. Two will be in direct conflict, and the third will produce the ultimate resolution. At sentences 18 and 19, where the detailed action of the story begins, the subjects brought together their tentative hypotheses and formulated general plans for the rest of the story. Table 5 shows some examples of what subjects said at this point. It was striking to us how regular this phenomenon was: virtually all subjects did the same thing at about the same place in the story.

Insert Table 5 about here

Throughout the processing of the main part of the story, the hypotheses constructed by the subjects were much in evidence. Each event was incorporated into the general plan constructed at the end of the sentence 17. The impact of these hypotheses was most dramatically revealed at sentences 30 and 31, where the climax occurred. The readers were certain that something was about to happen, but were uncertain as to exactly what. Consistent with the expectation that a story contains novel elements, the readers expected to be surprised by the specific form of the complication, even though they knew it would involve two particular characters. The first sentence of the climax is totally surprising, yet subjects immediately knew that the climax must be at hand and that the third character is about to intervene to save the day. This is all the more interesting because the third character has not been mentioned for 16 sentences. Subjects are led to expect that people are mentioned in a story for a reason. When a character is not brought in for a while, expectations of an appearance grows. The readers are anticipating a place for the character to fit in. Because the story is well written, at least by the conventions for children's stories, the predictions subjects make in response to the climactic event are in fact right.

Once the climax is passed, the mode of processing changed quite drastically. The readers appeared to be operating in a confirmatory mode. No new surprises were expected. Over and

over one finds in the protocols statements like "That's what I expected." or "Yup." indicating that predictions generated were now being found to be congruent with the emerging details of the post-climactic part of the story.

Lentil is a good example of a story where the writer and the reader are operating in harmony, and the TOL data give a clear picture of how this works. The early part of the story sets a background and leads to hypotheses which successfully guide the subsequent processing. The same expectations are derived by different readers at just about the same points in the story, and specific events tend to be interpreted in the same way. The climax is especially striking. Though it is unpredictable, it is immediately interpreted by all subjects and is integrated into the general representation being constructed by the reader.

In Olson et al. (1981) we discuss differences between the processing of Lentil and the ill-formed stories we investigated. These contrasts are especially informative in helping us construct the portrait of processing just presented. By seeing what subjects do when things are not working well, we obtain a clearer picture of the kinds of strategies subjects are trying to employ while reading.

Because they are written and read for different purposes, one would expect stories and argumentative essays to be read in somewhat different ways. An analysis of TOL data for stories and essays confirms this. Table 6 shows the relative frequencies of types of talking in the TOL data for two versions of two of the essays we used. The greater frequency of Comments on Structure

is due to our instructions: we added this to the set of things we stressed to our TOL subjects.

Insert Table 6 about here

The summary data in Table 6 do not reveal some important details regarding qualitative differences in the processing of stories and essays. Perhaps the most dramatic difference comes in the kinds of predictions made by readers. The typical predictions made by a reader of a story are specific. Subjects predicted events that might occur later in the story, involving specific characters. The examples in Table 5 are typical. The specificity of the predictions increased as the story developed, but even very early in the stories the predictions were remarkably specific. Those in Table 5 occurred less than a third of the way into Lentil. In marked contrast, the predictions given in the essays were much more general. They often consisted of comments to the effect that the reader expected to see another argument or another example. But the specific content of the argument or example was usually not predicted. This difference in the types of predictions generated by readers reflects some important differences in the way they approach the two text types. The reader of a story has a set of firm expectations about the type of substantive events that will constitute an acceptable continuation of the story. Given background information about the characters and their motives, much can be anticipated about how the conflict in the plot will arise and how it will be resolved. However, for a typical argument it is

apparently more difficult to predict the specifics. Readers know that an inductive argument will have various elements of supporting evidence for the main point, and they know that a compare and contrast essay will make comparisons and draw contrasts. But their expectations do not readily translate into specific predictions. For example, Table 7 gives typical predictions generated by readers at the end of the first paragraph of the well-formed version of the Carpeting essay (the complete essay is in Table 8; the first paragraph ends at sentence 6). This paragraph presents some background information, introduces the thesis of the essay, and announces that there will be three arguments made in support of the thesis. In contrast to the predictions given at the end of the background section in Lentil (see Table 3), the essay readers do not seem to go beyond what the author has told them about the arguments to be presented.

Insert Tables 7 and 8 about here

While the predictions in the essay TOL data were not as rich as those for stories, the comments on essay structure were extremely informative. These comments revealed subjects' expectations about how an essay should be written. For example, our readers expected to find a topic or thesis sentence, and they expected to find it early. In one version of our essay on the coming ice age, we deliberately placed the topic sentence late in the essay. As a result, pieces of evidence for the coming ice age were presented before the author announced the main point.

When reading this essay, many TOL subjects explicitly searched for the topic sentence and complained that the author took so long to come to the point (see Table 9). In the Carpeting essay, subjects revealed a sensitivity to surface signalling devices. For example, in response to a sentence that began with the phrase "In short," subjects predicted that the writer would now move on to the next point in the argument. When we presented subjects with a version of this essay with such signalling removed, we found evidence of confusion in the protocols.

Insert Table 9 about here

In short, the reader of an essay has general expectations about the overall structure of the argument or thesis. The reader quickly recognizes the type of point being made, and at a general level is sensitive to the organizing devices in the surface structure of the essay. However, unlike the story reader, the reader of an essay does not appear to engage in rich hypothesis generation and testing. The reader seems to adopt a more passive strategy, waiting for each new item of information to be presented and trying to fit it into the overall scheme of the argument.

A capsule summary of the differences in strategies for the readers of stories and essays might go as follows. The basic orientation of the reader of a story is prospective. The reader is looking ahead, trying to anticipate where the story is going. Except at the beginning, where an overall hypothesis is being developed, the story reader tends to relate each sentence to the

general hypotheses and predictions that have been developed. In contrast, the reader of the essay adopts a retrospective orientation. Each new element in the essay is related to earlier elements. There is little anticipation of what is coming up, except at the most general level. See Olson et al. (1980) and Olson et al. (1981) for a more detailed discussion of these differences and their basis in the nature of the two genres.

Quantitative analyses. An obvious question is whether the data obtained from a TOL task have anything to do at all with reading when not talking. There are certainly many peculiarities of the TOL situation that could distort the processes used by the reader, and thus give us a false impression of what is occurring during reading. There are many ways in which this could be assessed. In essence, one wants to see if properties of the TOL data in any way relate to properties of reading under other situations.

We have examined this by carrying out multiple regression analyses of TOL data, using properties of the TOL data as independent variables and sentence-by-sentence reading times as dependent variables. We measured reading times by having an independent group of subjects read each text at a computer terminal. Each time they pressed a key the next sentence of the text appeared. Only one sentence was shown at a time. Subjects were told to read the text as normally as possible. Those reading the stories were told that later we would explore how well they understood each story. Essay readers were told they would later have to write a one-sentence summary of each essay.

The primary data from this task are the times subjects devote to reading each sentence of each text.

Figure 2 shows the reading rates for each sentence for a group of 12 subjects reading Lentil. The measure of reading rate on the ordinate takes into account differences in sentence length. Though there are interesting connections between the qualitative picture of story processing revealed by our general analysis of the TOL data and the profile of reading times in Figure 2, here we want to focus on the quantitative analyses.

Insert Figure 2 about here

A stepwise multiple regression was performed, using mean reading time per sentence as the dependent variable and a series of independent variables. We focus on the stories first. For each story, two analyses were done, one that included sentence serial position and one that did not. Table 10 shows the general results (see Olson et al., 1981, for the specific regression coefficients). This table lists the predictors selected by the stepwise regression and the cumulative variance accounted for for each of the two analyses for the four stories. Note that for both Lentil and Stranger, our two well-formed stories, the relative frequency of various TOL categories accounted for significant portions of variance in the reading times, when the effect of sentence length has been removed as a separate factor. Note also that serial position and predictions are independent variables that are correlated with each other. When serial position is excluded from the analysis, predictions takes its

place as a predictor of reading times. The case is more mixed for the ill-formed stories. Only sentence length predicts reading times for Circle Island, while only one minor category of TOL productions accounts for any portion of the variance in reading times for War of the Ghosts. But the data for the well-formed stories are quite clear. Places where subjects in the TOL task generate more talking, especially predictions and inferences, are the same places where independent subjects slow down while reading silently. This supports the claim that the TOL data are related in an important way to what readers are doing during more ordinary types of reading.

Insert Table 10 about here

A quite different picture emerged for the essays. Multiple regressions of reading time data for the essays revealed no relationships between the TOL data and mean reading times. The only variable to emerge as a significant predictor was sentence length. This is in marked contrast to the story data, where several indices of TOL behavior correlated with the reading times. This is yet another indication of the differences in the behavior of the two groups of subjects.

One other difference between story and essay processing emerges in the quantitative analyses. Serial position was a predictor of reading times for the well-formed stories, but it does not emerge as a predictor for the essays. We assume that serial position is a proxy for predictive processing. When subjects adopt a predictive mode they read more slowly at the beginning of a text as they generate hypotheses, and more quickly later in the text because they are confirming earlier predictions. As we argued earlier, the TOL data provide evidence that subjects adopt this strategy when reading stories but not when reading essays. The presence of a serial position effect in the reading times for stories but not for essays provides additional evidence for this strategy difference between stories and essays.

Question-asking TOL for simple stories

One of the focussed TOL tasks we have used is one in which subjects ask questions following each sentence. We described the essence of this task earlier. In this section we present some data that show that the number of questions asked for each sentence correlates with sentence-by-sentence reading times.

Why use a question-asking task? We felt it tapped behavior relevant to what skilled readers do while reading. It seemed plausible to assume that each sentence encountered in a text raises certain questions in a reader's mind and answers other questions raised by earlier sentences. We wanted to explore this supposition in more detail by collecting rich data on the kinds of questions readers ask following each sentence in simple stories.

This study used four tasks. The primary task was one in which readers asked questions after reading each sentence in the story. In another task a different group of subjects read the same stories silently while we timed their reading. These same subjects later recalled the stories. Finally, another group of

subjects rated the importance of the constituents of the story. Four simple short stories (maximum length was 41 sentences) were used as texts. They were all children's stories or simple folktales, and all were well-formed.

To better understand the results, a somewhat more detailed description of the four tasks is necessary:

1. Question-asking. All four stories were presented to 9 subjects. Each sentence in the story was typed on a card, and the subject worked his or her way through the deck of cards, asking questions that were raised in his or her mind as a result of having read that particular sentence. The subject was told to imagine that the story's author was present, and that the author was willing to answer any questions the reader had about the story at that point, except for the obvious question of what happens next. The subject was allowed to spend as much time on any sentence as he or she desired, but was asked not to reread any previous sentences or to look ahead. The questions were tape recorded and later transcribed. The number of questions asked for each sentence was tallied and pooled over subjects. In addition, the questions were classified in various ways.

2. Reading times. Sentence-by-sentence reading times were collected from 20 subjects. At the end of each story subjects wrote a brief (3 to 5 sentences) summary of the story.

3. Recall. The same 20 subjects were asked to recall the stories they had just read. They were presented with a brief descriptive title for each story, and were given unlimited time to try to recall as much as they could. They were asked to

recall exact words, but were encouraged to guess if they could not remember exact words. Recall was scored by first doing a propositional analysis of each story and then matching the subject's recall against this, using a gist criterion.

4. Importance. Seventeen subjects read each story and crossed out the 50% of the words, phrases, or sentences in the story they felt was least important. For each sentence in each story the proportion of words left in averaged over subjects provided a measure of the relative importance of that sentence.

It is useful to have a better picture of what the question-asking data look like. Table 11 shows typical questions for the first sentence of one of the stories. These questions are grouped into those asked by two or more subjects and those that are idiosyncratic to one subject. Of course, we were also interested in the sentence-by-sentence variation in the questions asked. Figure 3 shows the total number of questions asked for each sentence in each of the four stories. With the possible exception of EMERALD, there is noteworthy variation in the number of questions asked from sentence to sentence. In EMERALD, there were a large number of questions at the beginning and then a fairly flat distribution of questions thereafter. Keep this difference in mind, because EMERALD will not follow the pattern of other stories in some of our later analyses.

Insert Table 11 and Figure 3 about here

The first issue we addressed was whether the question-asking task is related to the reading times. We examined this by

looking at the relationship between the total number of questions asked for each sentence in a story and the average reading time for each sentence for those subjects who were reading silently. The expectation was that sentences which elicited a lot of questions would be especially salient to real-time processing, and therefore would be read more slowly by subjects who were reading silently. This hypothesis was confirmed. We conducted a multiple regressions in which the average reading time per sentence was the dependent variable and several different predictor variables were explored. The predictor variables were sentence length, total number of questions, serial position, and importance. Only sentence length and number of questions emerged as significant predictors of reading time. In this analysis all four stories were entered, with story as a variable. There are two types of questions that occur: those that are asked by several subjects, and those that are idiosyncratic. We next asked whether these two types of questions contributed differentially to this outcome. The answer is no. A multiple regression with number of questions asked by two or more persons and idiosyncratic questions entered separately showed that both emerged as significant predictors. Table 12 shows the details of these analyses.

insert Table 12 about here

So, number of questions asked accounts for a significant portion of the variance in sentences-by-sentence reading times. We next asked what relationship the question-asking task has with

recall. And the answer was very simple: none. Table 13 shows the outcome of a multiple regression carried out on recall scores, and reveals that rated importance and serial position emerged as significant predictors of recall, while number of questions asked did not. This leads us to conclude that the information being revealed by the question-asking task is more closely associated with the activities that occur during comprehension than with the form of the final memory representation constructed as a result of comprehension.

insert Table 13 about here

This basic result confirms our initial supposition that the question-asking task would tap an aspect of what is going on in the skilled reader's mind while reading. The obvious question, of course, is what is it tapping? It is unlikely that a reader who is reading silently is actually asking questions while reading. Rather, we believe that the question-asking task taps the kinds of informational needs a reader encounters while proceeding through a text. As each sentence is understood and added to a growing representation of the story, the reader revises and elaborates the set of information still needed to have the developing story make sense. These informational needs interact with what is presented in the next sentence to generate a new set of informational needs--or, if you will, a new set of questions--that guide the reader's comprehension through the succeeding parts of the text.

We have conducted a number of other analyses of these data

that are discussed in Olson, Duffy, Eaton, Vincent, and Mack (in preparation). We have categorized the questions to see if certain types are more important than others. So far, the categories we have examined have not shown any differences. We have looked to see whether or not questions asked are later answered by the story, and there are interesting relationships here. Many questions are in fact answered, though it varies somewhat by type. However, the number of questions answered by a particular sentence does not predict reading time or recall. We have looked at the information tapped in the question, and find that questions which are derived from new information contained in the current sentence are especially important in predicting reading times. These and other details of these data are interesting and important, and will be reported on fully in Olson et al. (in preparation).

The main findings of this study strongly suggest that the question-asking task is a useful indicator of processes which may be an important part of comprehension. The number of questions asked by subjects as they read through a story correlates with the amount of time spent on that sentence by other readers reading silently. Keep in mind that this result is with the obvious effect of sentence length removed. But number of questions does not correlate with recall. Thus, question-asking seems more closely related to the real-time processes that occur during reading than to the final product of comprehension that remains when reading is completed.

This study is a nice example of the analytic usefulness of

the focussed TOL task. No claim is made that the question-asking task taps all or even many relevant aspects of comprehension. Rather, one particular theoretically promising component of comprehension processing is singled out for detailed treatment. As with the general TOL task, the analysis of these data can proceed in both a qualitative and a quantitative fashion.

4. RELATED APPLICATIONS OF TOL

In this chapter we have focussed on the use of TOL to reveal comprehension processes. TOL techniques are useful in some closely related domains, and in this section we present a few examples. Each of these uses TOL during comprehension either in a special environment or for a special purpose.

Computer Text Editing

The TOL method has been used to investigate how new (computer naive) users learn text-processing procedures with self-study instructional materials (see Lewis & Mack, 1982a, 1982b, 1982c; Mack, Lewis & Carroll, 1982). In this situation, the instructions were very general: new users were asked to talk about any aspect of their learning experience, including their interaction with the computer interface and the manual. They were asked to talk about any problems or questions they had, and any plans or decisions they might be aware of. Except for occasional non-directive prompts for reticent talkers, users decided when to talk and what to say. TOL data were augmented by a video-taped record of the subject working at the computer terminal.

The TOL data revealed much qualitative information about the

learning strategies and problems of new users. For example, self-study instructions were surprisingly "fragile" in that it was relatively easy for users to get side-tracked trying to follow them. This was due not only to simple oversights but also misunderstandings that reveal interesting reasoning strategies (see Lewis & Mack, 1982c; also Carroll & Mack, 1982a, 1982b). The new user's "innocence" about computers and their complexity made it surprisingly difficult for them to recover from these problems.

In this application of TOL Lewis et al. have not tried to relate qualitative observations to other more quantitative measures of immediate processing, although nothing would prevent doing so in principle. The qualitative data alone, however, have provided great insight into the problems of new users in a complex task domain. They have suggested a number of directions for more analytical investigation of learning, as well as practical applications in the design of interfaces and training methods. As such, it has already demonstrated the usefulness of TOL in research on an important genre of text: instructional materials in their real-world context of use.

TOL as Feedback to Writers

Recently, we have begun research that explores the usefulness of TOL data as feedback to writers. The rationale is quite simple. One of the difficulties that moderately skilled writers have is correctly discerning the state of mind of the reader. They, as writer, have the complete structure of their to-be-communicated ideas in mind. But it is difficult to imagine

the state of mind of the reader, who does not know these ideas and who may have a somewhat different general state of knowledge than the writer.

We discovered, somewhat accidentally, that TOL data provided marvelous feedback to a writer. We had prepared materials for various text comprehension experiments, and in some of our early pilot work on the TOL task we gave these texts to TOL subjects. The information we received from these subjects about what parts of our texts were hard to understand, which parts miscommunicated what we intended, and which parts violated the conventions of writer-reader communication was incredibly valuable. This led us to develop and use the selected TOL task in the preparation of stimulus materials. In addition, it suggested to us that TOL behavior might in general be a useful form of feedback to writers.

We are currently conducting research that directly examines this. Writers generate texts of various types from content we provide them, and then a series of readers provide process feedback about the sentence-by-sentence comprehension of these texts either by thinking out loud or by combining talking with doing in the case of texts that give instructions for how to do something. The writers are given a chance to revise their texts in light of the process feedback they receive from readers. Though we have just begun this work, our initial impressions are that this is an effective form of feedback to the writer.

We as investigators as well as the writers of textual materials can use TOL data as a measuring instrument for the

effectiveness of various texts. Indeed, since the effectiveness of particular texts is a joint function of text properties and reader properties, as Kintsch and his co-workers (Kintsch & Vipond, 1978; Kintsch & van Dijk, 1978) have so aptly demonstrated, studies using TOL tasks that varied both text properties and reader characteristics promise to provide especially informative data for developing theories of readers, writers, and textual transmission.

TOL and Metacognitive Awareness

Few people engage in TOL activities spontaneously. Does asking subjects to think out loud cause them to change their processing? The issue of whether or not such effects exist is central to the use of TOL methods (see Ericsson and Simon, 1980). Though such feedback effects from TOL are a methodological problem for the investigator interested in ordinary processing, they are a potential boon for the instructor who would like to improve the cognitive processing of a target population. Scardamalia and Bereiter (in press) have discussed this feature of TOL for children. There has been much research on the relationship between metacognitive awareness and comprehension with grade school children (see Brown, Bransford, Ferrara, & Campione, in press). Scardamalia and Bereiter have found informally that TOL methods often increase the metacognitive awareness of children. Though they have not yet conducted any formal studies of this relationship, their extensive experience with TOL methods suggests to them that it may be a significant instructional device for reading and writing with children.

5. WHY DO TOL DATA PRODUCE VARYING RESULTS?

In our discussions with colleagues about our TOL research we have heard of several disappointing efforts to use TOL to study comprehension. In fact, in our own work, our results for the essay TOL task were somewhat disappointing. We reported earlier the substantial differences in both the richness of the TOL protocols and in the nature of their correlational relationship with sentence-by-sentence reading times for simple stories and essays. These mixed results are scarcely surprising. Any empirical technique will have successful and unsuccessful applications. When will TOL be useful? In considering some of the reports of disappointing outcomes and our own successful applications of the techniques, we have come up with several factors that can affect how useful the TOL technique will be.

Types of instructions. We stressed earlier in this chapter the importance of being clear and explicit to subjects. The antecedent to this, of course, is understanding precisely what it is that one wants to get out of the task. Different instrucrions will produce quite different outcomes.

It is also important to make sure the instructions are appropriate to the texts being used. We have speculated that our essay TOL data were disappointing in part because of the instructions we gave subjects. Recall that our instructions for the essay TOL task were almost identical to those for stories. We used similar instructions so we could compare results across genres. The instructions, however, might have been inappropriate for essays. For example, we asked subjects to make predictions,

yet the protocols revealed that making rich content predictions was an inappropriate task for reading essays. A more appropriate task might have been to ask subjects to evaluate how convincing an argument was. In fact some subjects spontaneously adopted a more critical mode of talking about the argumentative essays. In these protocols the talking seemed more natural.

Subjects. Even with clear instructions, not all subjects will talk equally informatively. We have found that some of the best subjects in our research have been faculty and graduate students in psychology, who have at least a passing acquaintance with TOL methods and therefore know the level of information we are seeking. Some subjects do not know this, even with explicit instructions. Where large pools of appropriate subjects are not available, training subjects to talk may be a way of ensuring reasonable quality data. The exact content of what subjects say has to be up to them, of course. But the amount and level of talking can be inappropriate and may be subject to training.

We have also found that large individual differences exist in how subjects read some texts. This seems especially true for the essay TOL task. Some subjects adopted a critical mode in reading the argumentative essays. Other subjects did not. Some subjects talked easily about aspects of essay structure (e.g., topic sentences, conclusions, pro and con arguments). Others did not. We assume these differences in talking reflect differences in strategies readers adopt when reading essays or differences in the knowledge readers have about the genre. While these individual differences might be interesting in themselves, they

made it difficult to find common patterns in the TOL data for the essays. This heterogeneity was not so apparent in the story protocols where subjects seemed to have a common approach and knowledge base to use in reading the stories.

Type of material. One major problem for the researcher in text comprehension is to find or construct appropriate texts. Too often investigators in this area have used "stories" that are not really stories or paragraphs that are so artificial they do not resemble naturally occurring paragraphs. Our impression is that the richest protocols are elicited by texts that are natural and interesting. Our two well-formed stories were real children's stories. The plots were engaging enough to motivate the reader to read on to find out what happens. The TOL protocols were correspondingly rich. In contrast, the essays we used were, frankly, rather bland and boring. Subjects had no intrinsic reason to want to keep reading. The resulting TOL protocols were also rather boring. While the differences in our essay and story protocols may be in part due to genre differences, we do not believe that is the whole story. We suspect that essays with more controversial or interesting content might have elicited richer TOL data.

When TOL data are used in the context of discovery, it is especially useful to include a variety of text types in the set of stimuli. Our strategy of using well-formed and ill-formed texts, or of comparing different versions of texts (e.g., the essays) is a useful way of validating the quality of the data being obtained. If the TOL data are the same for well-formed and

ill-formed texts, the investigator should be suspicious of what the subjects are doing.

What is analyzed. There are many different ways to analyze data as rich as those obtained from TOL tasks. Whether one gets useful information or not will depend upon what one looks for. For instance, our essay TOL data yielded some general, useful information about the overall strategies used in reading the essays. But the multiple regressions exploring the ties between TOL behavior and reading times for essays did not yield much. We are currently carrying out a number of other more detailed analysis of our TOL data. For instance, we are in the midst of a detailed content analysis of our original story TOL protocols, coding the chains of hypotheses and other interconnections in the data as a possible clue to further aspects of the reading strategies of subjects. Similarly, we have conducted a number of other analyses of our question-asking data that also get at further aspects of the representation of comprehension processes in reading (Olson et al., in preparation). Not all analyses we have attempted have panned out. It will require broader experience with the use of TOL data to study comprehension before we will have a clearer picture of the types of analyses that are generally more useful.

6. SUMMARY AND CONCLUSIONS

We have used the TOL task as one window into the reader's mind. In this chapter we have described various versions of the tasks, discussed the overall rationale for the technique and discussed its application to various domains. We will now

summarize by presenting in concise form a list of advantages and a list of limitations in using the TOL task to study the process of comprehension.

Advantages

1. The primary goal in using the TOL task is to explicate the higher level processes involved in comprehension. When used appropriately, it indeed seems to do this. In fact, it may be one of the few techniques available for getting at this level of comprehension activity.

2. TOL behavior, under at least some situations, appears to correlate with other forms of reading behavior, such as sentence-by-sentence reading times.

3. Though we have not done this yet in our research, TOL data in general have proven to be a useful means for studying individual differences in higher level cognitive processes (e.g., Newell & Simon, 1972). Studies of readers of varying levels of skill or varying degrees of background knowledge could profitably be pursued with this method, though there might be some difficulty in a confounding between level of reading skill and ability as a TOL subject (see below).

Limitations

1. The TOL task is sensitive to instructional variables. The instructions must be precise and must be carefully thought out in relation to one's research goals. Vague or very general instructions in general do not work well. Further, in light of Ericsson and Simon's (1980) analysis, it is important that the task focus on the reporting of current states of knowledge and

not ask subjects to report on states of knowledge very far in the past or to offer explanations for their behavior.

2. The TOL task seems to work better for some text types than for others. In our research, we felt we obtained much better TOL data for stories than for essays. However, so far the task has not been used for a very wide range of text types. Further, as we suggested earlier, there are undoubtedly important interactions between text type and instruction that are not very well documented yet.

3. TOL data are difficult to analyze. Any form of data collection which monitors a continuous stream of behavior over long intervals of time produces data which can be difficult to analyze. The transcription, coding, and analysis of TOL protocols is extremely time-consuming, and little of it can be automated. Thus, the decision to use the TOL task must be thoughtful and must take into account cost-benefit ratios.

4. There appear to be big differences among subjects in their ability to provide informative TOL data. Some subjects are good talkers, some are not. The difficulty is one of getting talking at an appropriate level and in appropriate quantity to be useful to the investigator. This problem can be met by training subjects to talk, but that is time-consuming. There are probably limits in how young or how intelligent TOL subjects can be.

5. The TOL task may influence the nature of the comprehension processes used by subjects. The fact that we get correlations between properties of TOL data and silent reading times suggests that similar things may be going on in the two

situations. But having subjects think out loud certainly has the possibility for distorting their processing, and it is essential that TOL data always be used along with other converging evidence in order to determine what it is that readers are doing as they work their way through a text.

References

Bartlett, F.C. Remembering: An experimental and social study. Cambridge: Cambridge University Press, 1932.

Brown, A.L., Bransford, J.D., Ferrara, R.A., & Campione, J.C. Learning, remembering, and understanding. In J.H. Flavell & E.M. Markman (Eds.), Carmichael's manual of child psychology (Vol. 1). New York: Wiley, in press.

Carroll, J.M., & Mack, R. Learning to use a word processor: By doing, thinking and knowing. In J. Thomas & M. Schneider (Eds.), Human factors in computer systems. Norwood, N.J.: Ablex, in press. (a)

Carroll, J.M., & Mack, R. Actively learning to use a word processor. In W. Cooper (Ed.), Cognitive aspects of skilled typewriting. New York: Springer-Verlag, in press. (b)

Collins, A., Brown, J.S., & Larkin, K.M. Inference in text understanding. In R.J. Spiro, B.C. Bruce, & W.F. Brewer (Eds.), Theoretical issues in reading comprehension. Hillsdale, N.J.: Erlbaum, 1980.

Dawes, R. Memory and the distortion of meaningful written material. British Journal of Psychology, 1966, 57, 77-86.

Ericsson, K.A., & Simon, H.A. Verbal reports as data. Psychological Review, 1980, 87, 215-251.

Graesser, A.C. Prose comprehension beyond the word. New York: Springer-Verlag, 1981.

Kintsch, W., & van Dijk, T.A. Toward a model of text comprehension and production. Psychological Review, 1978, 85, 363-394.

Kintsch, W., & Vipond, D. Reading comprehension and readability in educational practice and psychological theory. In L.W. Nilsson (Ed.), Memory: Processes and problems. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Lewis, C. Using thinking aloud protocols to study the "cognitive interface." Research Report RC 9265, IBM Thomas Watson Research Center, Yorktown Heights, New York, 1982.

Lewis, C., & Mack, R. The role of abduction in learning to use text-processing systems. Paper presented at the annual meeting of the American Educational Research Association, New York, March 1982. (a)

Lewis, C., & Mack, R. Learning to use a text processing system: Evidence from "thinking aloud" protocols. Paper presented at the Conference on Human Factors in Computer Systems, Gaithersburg, Md., March 1982. (b)

Lewis, C., & Mack, R. Why is it hard to write a good instruction manual? Ms., 1982. (c)

Mack, R., Lewis, C., & Carroll, J. Learning to use word processors: Problems and prospects. Research Report RC 9712, IBM Thomas Watson Research Center, Yorktown Heights, New York, 1982.

Neisser, U. Cognitive psychology. New York: Appleton-Century-Crofts, 1967.

Newell, A., & Simon, H.A. Human problem solving. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Nisbett, R.E., & Wilson, T.D. Telling more than we can know: Verbal reports on mental processes. Psychological Review,

1977, 84, 231-259.

Olson, G.M., Duffy, S.A., Eaton, M.E., Vincent, P., & Mack, R.L. On-line question-asking as a component of story comprehension. Manuscript in preparation.

Olson, G.M., Duffy, S.A., & Mack, R.L. Applying knowledge of writing conventions to prose comprehension and composition. In W.E. McKeachie (Ed.), Learning, cognition, and college teaching. San Francisco: Jossey-Bass, 1980.

Olson, G.M., Mack, R.L., & Duffy, S.A. Cognitive aspects of genre. Poetics, 1981, 10, 283-315.

Rumelhart, D.E. Understanding understanding. Center for Human Information Processing Report No. 100, University of California at San Diego, January 1981.

Scardamalia, M., & Bereiter, C. Child as co-investigator: Helping children gain insight into their own mental processes. In S. Paris, G.M. Olson, & H.W. Stevenson (Eds.), Learning and motivation in the classroom. Hillsdale, N.J.: Erlbaum, in press.

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

Footnotes

'This research has been supported by a research grant from the National Institute of Education (NIE G-79-0133) and by a Research Career Development Award from NICHD (HD 00169) awarded to the first author. The authors are grateful for the comments of Mary Eaton, Mike Fehling, Judy Reitman, Lynn Roshwalb, and Mary Trahan on earlier drafts of this chapter. Communications about this work should be addressed to Gary M. Olson, Department of Psychology, University of Michigan, 330 Packard Road, Ann Arbor, Michigan 48104.

Table 1

Levels of Analysis in Reading

PERCEPTUAL ANALYSIS
Perception of Features
Perception of Letters
Perception of Letter Patterns
DEVELOPMENT OF WITHIN-SENTENCE REPRESENTATION
Recognition of Words
Syntactic Analysis
Semantic Analysis
Pragmatic Analysis
DEVELOPMENT OF INTEGRATED REPRESENTATION ACROSS SENTENCES
Inferences
Elaboration
Hypotheses

Table 2

Sample Thinking-Out-Loud Protocol of a Subject reading Lentil

1. IN THE TOWN OF ALTO, ONIO, THERE LIVED A BOY NAMED LENTIL.

   In Alto, Ohio, is sort of . . . I assume this is going to be a story . . . something about music . . . or vegetables because we have the boy named Lentil.

2. LENTIL HAD A HAPPY LIFE EXCEPT FOR ONE THING.

   Most of this story is going to be developed around this one exception. He is generally happy, except for this one thing.

3. HE WANTED TO SING -- BUT HE COULD NOT.

   Ah hah. I'm remembering 'Alto, Ohio', that it is going to be about music, and he wanted to sing is the one thing that was unhappy in his life. The story is going to be about how he learns to sing. He is either going to go out and find birds and learn from them, or he is going to have a trumpet like the trumpeter swan. The story is going to be about fixing the fact that he can't sing.

4. IT WAS MOST EMBARRASSING, BECAUSE WHEN HE OPENED HIS MOUTH TO TRY, ONLY STRANGE SOUNDS CAME OUT.

   I don't see what . . . it's still talking about the fact that he can't sing (repeats) . . . no, we are just elaborating on the fact that he can't sing.

5. AND HE COULD NOT EVEN WHISTLE BECAUSE HE COULDN'T PUCKER HIS LIPS.

   More elaboration about his unhappiness . . . he can't produce music at all . . . 'He couldn't even whistle.' Now I'm beginning to think about how we are going to resolve this. Is it going to have anything to do

with the fact that his name is Lentil? or vegetables? Is he going to do anything with vegetables?

6. BUT HE DID WANT TO MAKE MUSIC, SO HE SAVED UP ENOUGH PENNIES TO BUY A HARMONICA.

That seems reasonable. He is going to try different ways to make music. He wants to make music, not just sing.

7. LENTIL WAS PROUD OF HIS NEW HARMONICA, AND HE DECIDED TO BECOME AN EXPERT.

So now he's going to learn it from somebody else.

8. SO HE PLAYED A LOT, WHENEVER AND WHEREVER HE COULD.

Now I expect this to be generating a lot of annoyance from his mother . . . having noisy children. (Repeats)

9. HIS FAVORITE PLACE TO PRACTICE WAS IN THE BATHTUB, BECAUSE THERE THE TONE WAS IMPROVED ONE HUNDRED PERCENT.

Like singing in the shower. (repeats) How we are going to have developments about him turning into a prune or something . . . a prune skin. (repeats)

10. HE USED TO PLAY ALMOST ALL THE WAY TO SCHOOL.

Now that's a place not in the bathtub. He "used to play" means he doesn't anymore. Maybe the children made fun of him, or something like that.

11. DOWN VINE STREET TO THE CORNER OF MAIN, PAST THE FINEST HOUSE IN ALTO, WHICH BELONGED TO THE GREAT COLONEL CARTER.

This isn't even a sentence. He is playing almost all the way to school. This is the route. Colonel Carter is just been introduced, and he is going to develop something with this little Lentil boy.

"Used to play almost all the way to school." We haven't done anything with the bathtub. It might be . . . this sounds like a children's story. So that's really reading about children making noise in the bathtub. Children love the sound of their own voices. Yell in parking structures and tunnels and stuff like that. That may develop, may not.

12. THEN PAST THE DRUG STORE, THE BARBER SHOP, AND THE ALTO LIBRARY, WHICH WAS A GIFT OF THE GREAT COLONEL CARTER, BY THE METHODIST CHURCH, THROUGH THE CARTER MEMORIAL PARK, AND AROUND THE SOLDIERS AND SAILORS MONUMENT THAT THE COLONEL HAD BUILT THERE.

So now we know Colonel Carter is rich . . . not only famous, but rich. "Then by the methodist church" . . . well, you could interpret this . . . figure that the Alto library is a gift of the great Colonel Carter. We are describing more of his wealth. "He goes by the methodist church, through the memorial park". More of Carter's stuff . . . very famous . . . and "around the soldier's and sailor's monuments, which the Colonel had built there" . . . sort of a central figure now. He is going to develop in some fashion.

13. THEN LENTIL WOULD STUFF HIS HARMONICA INTO HIS POCKET AND TAKE A SHORT CUT UP THE ALLEY BEHIND THE HARDWARE STORE SO HE WOULD NOT BE LATE FOR SCHOOL.

Why would he . . . apparently all this stuff -- playing -- slows him down and he has to hurry up to get to school. "So he stuffs it in his pocket and takes a short cut up the alley behind the hardware store, so he wouldn't be late for school."

14. PEOPLE WOULD SMILE AND WAVE HELLO TO LENTIL AS HE WALKED DOWN THE STREET, BECAUSE EVERYONE IN ALTO LIKED LENTIL'S MUSIC -- THAT IS, EVERYONE BUT OLD SNEEP.

I'll bet you Old Sneep is Colonel Carter. That's got to be his nickname. He's going to be . . . he is going to put an end to Lentil's playing. This is the story of a little boy against

---

the giant . . . David and Goliath. He is going to win over Old Sneep in the end.

15. OLD SNEEP DID NOT LIKE MUCH OF ANYTHING OR ANYBODY.

Yep, we are developing Old Sneep, which I'm guessing is Colonel Carter.

16. HE JUST SAT ON A PARK BENCH AND WHITTLED AND GRUMBLED.

Well, maybe it may not be. Can't imagine an old wealthy person sitting on a park bench, whittling and grumbling. But it could be all the same person.

17. ONE DAY THE NEWS GOT AROUND THAT THE GREAT COLONEL CARTER, WHO HAD BEEN AWAY FOR TWO YEARS, WAS COMING HOME.

Well, now I'm deciding that Old Sneep is not the great Colonel Carter. So we have Old Sneep, little Lentil who is playing all the way to school, through, across, in front of Colonel Carter's house, and Colonel Carter, who is great, rich, and magnificent and all that. Haven't decided whether Colonel Carter is a good guy or a bad guy.

18. PEOPLE BEGAN TO PLAN A GRAND WELCOME.

That's a welcome home for Colonel Carter. Oh, maybe we're going to have a parade and Lentil gets to be first in the parade or something like that.

19. BUT WHEN OLD SNEEP HEARD THE NEWS HE SAID, "NUMPH. WE WUZ BOYS TOGETHER. HE AIN'T A MITE BETTER'N YOU OR ME AND HE NEEDS TAKIN' DOWN A PEG OR TWO."

All right, now I know that Old Sneep is not Colonel Carter. Maybe not . . . maybe . . . maybe not . . . probably not. So we have a humbug here, who is unhappy. So he is going to try to destroy the parade, or whatever we're going to do -- the grand welcome.

Table 3

The Well-Formed Story Lentil

1. In the town of Alto, Ohio, there lived a boy named Lentil.
2. Lentil had a happy life except for one thing.
3. He wanted to sing - but he couldn't.
4. It was most embarrassing, because when he opened his mouth to try, only strange sounds came out.
5. And he couldn't even whistle because he couldn't pucker his lips.
6. But he did want to make music, so he saved up enough pennies to buy a harmonica.
7. Lentil was proud of his new harmonica, and he decided to become an expert.
8. So he played a lot, whenever and wherever he could.
9. His favorite place to practice was in the bathtub, because there the tone was improved one hundred percent.
10. He used to play almost all the way to school.
11. Down Vine Street to the corner of Main, past the finest house in Alto, which belonged to the great Colonel Carter,
12. Then past the drugstore, the barber shop, and the Alto Library, which was a gift of the great Colonel Carter, by the Methodist Church, through the Carter Memorial Park, and around the Soldiers and Sailors Monument that the Colonel had built there.

13. Then Lentil would stuff his harmonica into his pocket and take a shortcut up the alley behind the hardware store so he would not be late for school.

14. People would smile and wave hallo to Lentil as he walked down the street, because everyone in Alto liked Lentil's music - that is, everybody but Old Sneep.

15. Old Sneep didn't like much of anything or anybody.

16. He just sat on a park bench and whittled and grumbled.

17. One day the news got around that the great Colonel Carter, who had been away for two years, was coming home.

18. People began to plan a grand welcome.

19. But when Old Sneep heard the news he said, "Humph! We wuz boys together - he ain't a mite better'n you or me and he needs takin' down a peg or two."

20. Sneep just kept right on whittling, but everybody else kept right on planning.

21. Colonel Carter was the town's most important citizen, so the people hung out flags and decorated the streets.

22. The mayor prepared a speech.

23. The Alto Brass Band put on their new uniforms.

24. And the printer, the grocer, the plumber, the minister, the barber, the druggist, the ice man, the school teachers, the housewives and their husbands and their children - yes, the whole town went to the station to welcome Colonel Carter.

25. The train pulled in.

---

26. The musicians in the band were waiting for the leader to signal them to play.

27. The leader was waiting for the mayor to nod to him to start the band.

28. And the mayor was waiting for Colonel Carter to step from his private car.

29. All the people held their breath and waited.

30. Then there was a wet sound from above.

31. Slurp! There was Old Sneep, sucking on a lemon.

32. Old Sneep knew that when the musicians looked at him their mouths would pucker up so they could not play their horns.

33. The whole band looked up at Old Sneep.

34. The mayor gave the signal to play, but the cornetist couldn't play his cornet, the piccolo player couldn't play his piccolo, the trombone player couldn't play his trombone, and the tuba player couldn't play his tuba, because their lips were all puckered up.

35. They couldn't play a single note!

36. The musicians just stood there holding their instruments and looking up at Sneep sucking on the lemon.

37. The leader looked helpless.

38. The people were too surprised to move or say a thing.

39. And the mayor wrung his hands and wore a look that said: "Can't somebody do something, please?"

40. As Colonel Carter stepped from his car, the only sound was the noise of Sneep's lemon.

---

41. Clouds began to gather on the colonel's brow and he said, "Humph!" in an indignant sort of way.

42. Of course Lentil's lips were not puckered and he knew something had to be done.

43. So he took out his harmonica and started to play "Comin' 'Round the Mountain When She Comes."

44. When Lentil began to play the second chorus, Colonel Carter smiled.

45. Then he let out a loud chuckle and began to sing, "... driving six white horses when she comes."

46. Then everybody sang and they all marched down Main Street behind the colonel's car.

47. Lentil rode with the colonel, who took a turn at the harmonica when Lentil's wind began to give out.

48. (He said he hadn't played one since he was a boy, but he did very well considering.)

49. They marched to the colonel's house and paraded through the gate and onto the front lawn.

50. The mayor's committee served ice cream cones to all the citizens and Colonel Carter made a speech saying how happy he was to be home again.

51. When he said that he was going to build a new hospital for the town of Alto, everybody was happy -- even Old Sneep!

52. So, you never can tell what will happen when you learn to play the harmonica.

Table 4

Proportion of Thinking-Out-Loud Productions in Each Category for Stories

Category	Lentil	Stranger	Ghosts	Circle Island
Predictions	.22	.23	.13	.26
Questions	.04	.02	.10	.01
Comments on structure	.09	.06	.08	.10
Comments on own behavior	.03	.03	.04	.02
Confirmation of predictions	.02	.03	.01	.02
Referehces to antecedent information	.29	.36	.30	.27
Inferences	.30	.24	.30	.27
General knowledge and associations	.02	.04	.03	.06

Table 6

Examples of Reader Comments to Sentence 18 in "Lentil"

Subject	Comments
24	"I expect the plot will succeed in getting the Colonel to hear Lentil playing his harmonica. The Colonel will be impressed and Lentil will be rewarded somehow."
17	"Maybe a celebration is planned (parade, etc.) and Lentil will win the day with a rousing welcoming song."
07	"Expect to read that Colonel Carter now will have some kind of role in what's going on with Ella and his music and Old Sneep. Expect to hear something about Colonel Carter's reaction to Lentil."
23	"We expect to see some interaction between Colonel Carter, Old Sneep, and Lentil. Probably a great Celebration. Suspect that Lentil will probably be asked to play for him."

Table 6

Proportion of Thinking-Out-Loud production in Each Category for Essays

Category	Ice Ages well-formed	Ice Ages ill-formed	Carpet well-formed	Carpet ill-formed
Predictions	24	29	18	24
Questions	12	12	06	08
Comments on structure	32	28	28	32
Comments on own behavior	03	04	03	03
Confirmation of predictions	01	02	01	03
References to antecedent information	18	14	28	19
Inferences	10	12	11	10
General knowledge and associations	01	09	09	01

Table 8

Well-formed Version of Carpeting Essay

1. I should start by admitting that as little as five years ago carpeted classrooms would rightly have seemed too fanciful and expensive.

2. The carpeting then available would have been costly, difficult to maintain, and would have required frequent replacement.

3. Now, however, because of improved materials the arguments in favor of extensive carpeting seem a great deal more plausible.

4. New indoor-outdoor synthetics are stain resistant, fade resistant, durable and inexpensive.

5. They have made carpeting seem much less a luxury than a reasonable, even desirable, alternative to tile floors.

6. Briefly, there seems to be three central arguments in favor of carpeting.

7. First, of course, carpeting is attractive.

8. Now, admittedly, modern technology offers a great variety of attractively colored tiles.

9. The days of dreb, institutional greys, greens, and browns in tile are over.

10. But while tile may approach carpeting in terms of color it has a hard and unattractive texture.

11. Carpeting, on the other hand, is colorful, attractive to the touch, and comfortable to walk on.

12. It goes a long way toward creating a pleasant atmosphere all of us would like to work in, both in and out of class.

13. Richly colored carpeting, such as bold reds often used in banks and commercial offices, would make our facilities less institutional.

14. Bright carpeting can easily make attractive an area that would otherwise seem Spartan and sterile.

15. In short, carpeting seems desirable simply because it is more attractive to look at and walk on than tile.

Table 7

Examples of Reader Comments at the End of the Introductory Paragraph of the Well-formed Carpeting Essay

Subject	Comments
02	"It's getting to go on to describe what is good about carpeting. Before it just told how it came to be that they're using more carpeting. I think they're going to tell the reasons why."
03	"I expect to find out what the three central arguments in favor of carpeting are."
04	"Now you would expect him to go on and possibly list and possibly explain each of the arguments in favor of carpeting."
06	"Sounds like they are going to sell me carpeting now. They will go into all the good points of carpeting and try to sell me a roll."

16. The second argument in favor of carpeted classrooms is essentially that carpeting serves a useful aroustical function.

17. Of course, the flexible backing and rough texture of modern tiles make them far less noisy than those of just a few years ago.

18. Both tiles and carpets have improved significantly.

19. Carpeting, however, is a superior dampenar of sound.

20. It cuts noise from crowded hallways, absorbs annoying background noise in classrooms and makes busy space less noisy, and therefore more practical.

21. In industry, if not in schools, one frequently finds carpeting in busy areas because it reduces noise.

22. A final argument in favor of carpeting is that over a period of time, carpeting is no more expensive than tile.

23. Certainly, carpeting cost more than tile and it does need eventual replacement.

24. But carpeting costs much less to maintain than tile, which needs frequent washing, waxing and dusting.

25. The new synthetic carpets resist stains and fading.

26. An ordinary vacuum cleaner will keep them in shape.

27. But the tile floor, unfortunately, needs frequent scrubbing and waxing, if it is not to look dull and yellow with accumulated wax.

28. This process is laborious and slow, and, in large institutions, it requires expensive scrubbing machines.

29. In short, tile costs less than carpeting to install.

30. But count in the maintenance, and carpeting becomes a legitimate economic alternative to tile.

31. Were it not for the advantages in appearance and acoustics of carpeting, one could perhaps argue in favor of conventional tile flooring.

32. After all, the costs over a very long period, say twenty or thirty years, are genuinely unpredictable.

33. We simply have not accumulated enough experience with the new synthetics.

34. Perhaps over a quarter of a century carpets will prove more expensive.

35. Perhaps we will discover that after a decade or so, the savings in maintaining carpets will evaporate.

36. To this point, however, our expetience with the new synthetic materials is essentially affirmative.

37. And so, given the clear edge carpeting has over tile aesthetically and acoustically, and its economic justificstion, carpeting seems sensible.

Table 9

Comments from Thinking-Out-Loud Subjects Reading

Late-Thesis Version of Ice Age

Sentence	Subject 6
4.	With this sentence he is setting up what the subject will actually be. And the next sentence will almost certainly be what his essay is actually going to be about.
5.	Instead of telling right out what the problem is, the person is delaying it a little bit, and now you get a sense of unease -- wondering what it is this person is talking about. I would expect though that the subject, the main subject of this essay will be coming up very shortly.
12.	Again there are more hints about ice ages here; even though the ice age -- he has not really said anything about it. I now believe that the whole structure of the essay up to this point has been to keep the reader uneasy, just dropping little bits of information until finally he is aware that we should be expecting another ice age even though he has never said so yet.
13.	OK, finally he does say that we're on the verge of another ice age.
	Subject 8
4.	I expect the next sentence to tell us exactly what this serious problem is.
5.	This is the beginning of a new paragraph, and I really don't know what they're going to be talking about yet. I think that perhaps the serious problem that the essay's going to discuss should have been mentioned in the first paragraph or somehow it's going to be tied in in the next couple of sentences.
6.	That's fine, but what does it have to do with a serious problem?
7.	It looks like they're giving us all the symptoms of a serious problem, but we don't really know what it is.
8.	Wonderful, we still don't know why we're discussing this.

9.	Somehow I feel I missed the whole sentence tying this together. We still haven't been told exactly what the essay's going to be about, the main theme. All that's been done is examples after examples.
10.	All these examples are fine. However we're halfway done with the essay and we still don't know what we're talking about.
13.	This sounds like the theme of our essay, now that we're halfway through. I think the introduction was rather long.
	Subject 9
4.	I expect him now to give me what this more serious problem is.
11.	It has taken me 11 sentences to figure out where this essay is going to go. It doesn't really seem that those 11 sentences have done a good job of telling me what the point of the essay is going to be.
13.	This sentence could have started the entire essay.

Note: Sentence 4 states: "But mankind may soon be facing a more serious problem than any of these."
Sentence 13 is the thesis statement: "It now seems probable, climatologists say, that the world is on the verge of another ice age."

Table 10

Significant Predictors of Sentence-by-Sentence Reading Times for Simple Stories

	Lentil		Stranger		Ghosts		Circle Island	
Analysis with Serial Position	Syll.	.696	Syll.	.493	Syll.	.743	Syll.	.604
	Ser.Pos.	.748	Ser.Pos.	.588	Gen.Know.	.826		
	Inf.	.789	Gen.Know.	.659				
Analysis without Serial Position	Syll.	.696	Syll.	.493	Syll.	.743	Syll.	.604
	Inf.	.742	Imp.	.540	Gen.Know.	.826		
	Pred.	.762	Pred.	.603				
			Gen.Know.	.666				

Note: Syll.=number of syllables, Ser.Pos.=serial position, Gen.Know.=general knowledge, Inf.=number of inferences, Pred.=number of predictions, Imp.=rated importance. Table entries are the cumulative variance accounted for by a stepwise multiple regression, listed in each cell in the order they were selected.

Table 11

Sample Questions from The Selling of the Cow

Questions asked by 2 or more subjects:	# Subjects
1) Who is Cromer?	2
2) What is Cromer like?	3
3) Did Cromer live alone?	5
4) When did this story take place?	5
5) Where was the farm?	4
6) Where was the hill?	4
7) Why was the farm on a hill?	2
8) How far up the hill was the farm?	2
9) How high was the hill?	3
10) What kind of farm was it?	4
11) What will happen to Cromer?	2
Idiosyncratic questions asked by only 1 subject:	
1) Does the fact that he lives on a farm have any significance?	
2) Does he farm for a living?	
3) Does he have another vocation?	
4) Is Cromer married?	
5) How old is Cromer?	
6) How far away were Cromer's nearest neighbors?	
7) Why did Cromer like to live on a farm?	

8) Are they going to roll something down the hill?

9) Did a lot of the dirt wash off the side of the hill so that Cromer couldn't have his crops?

10) What was Cromer's first name?

11) Was that Cromer's first name?

12) Then what was Cromer's last name?

13) Did Cromer have more than one name?

14) What kind of name is Cromer?

15) What does Cromer mean?

16) What nationality is Cromer?

Table 12

Multiple Regression Analyses of Reading Time in Question-Asking Experiment

	Coeff.	Sig.	Cum. R^2
1. Predictors selected:			
Sentence length	130.22	.0001	.589
Total number of questions asked	26.44	.0001	.640
Predictors not selected:			
Serial position of sentence			
Importance			
2. Predictors selected:			
Sentence length	130.81	.0001	.589
Idiosyncratic questions	40.27	.0004	.626
Number of questions asked by two or more subjects	87.42	.0015	.652
Predictors not selected:			
Serial position of sentence			
Importance			

Note: Forward stepwise regression, dependent variable=mean reading time per sentence.

Table 13

Multiple Regression Analyses of Recall in Question-Asking Experiment

	Coeff.	Sig.	Cum. R^2
Predictors selected:			
Importence	341.12	.0001	.235
Serial position of sentence	-2.94	.0024	.283
Predictors not selected:			
Sentence length			
Total number of questions esked			

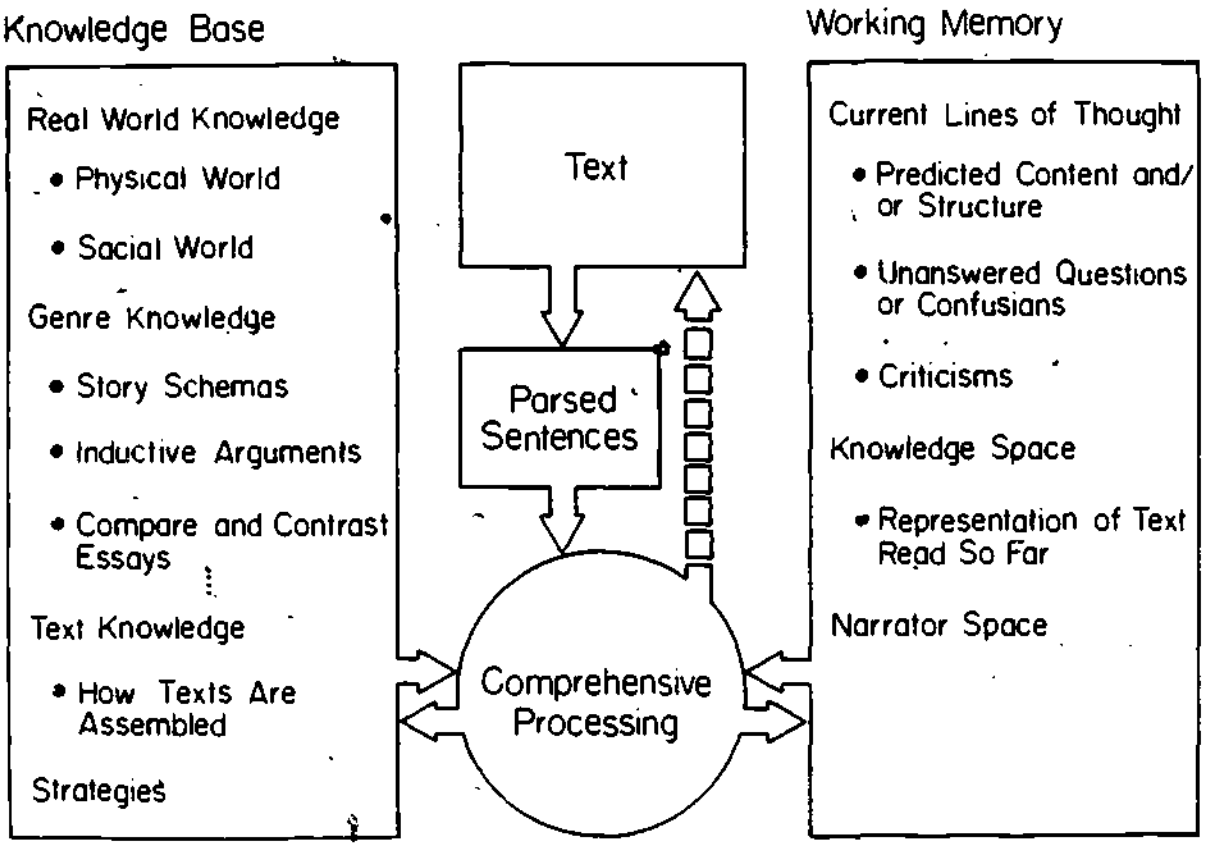
Note: Forward stepwise regressions, dependent variable=proportion propositions recalled per sentsnce.

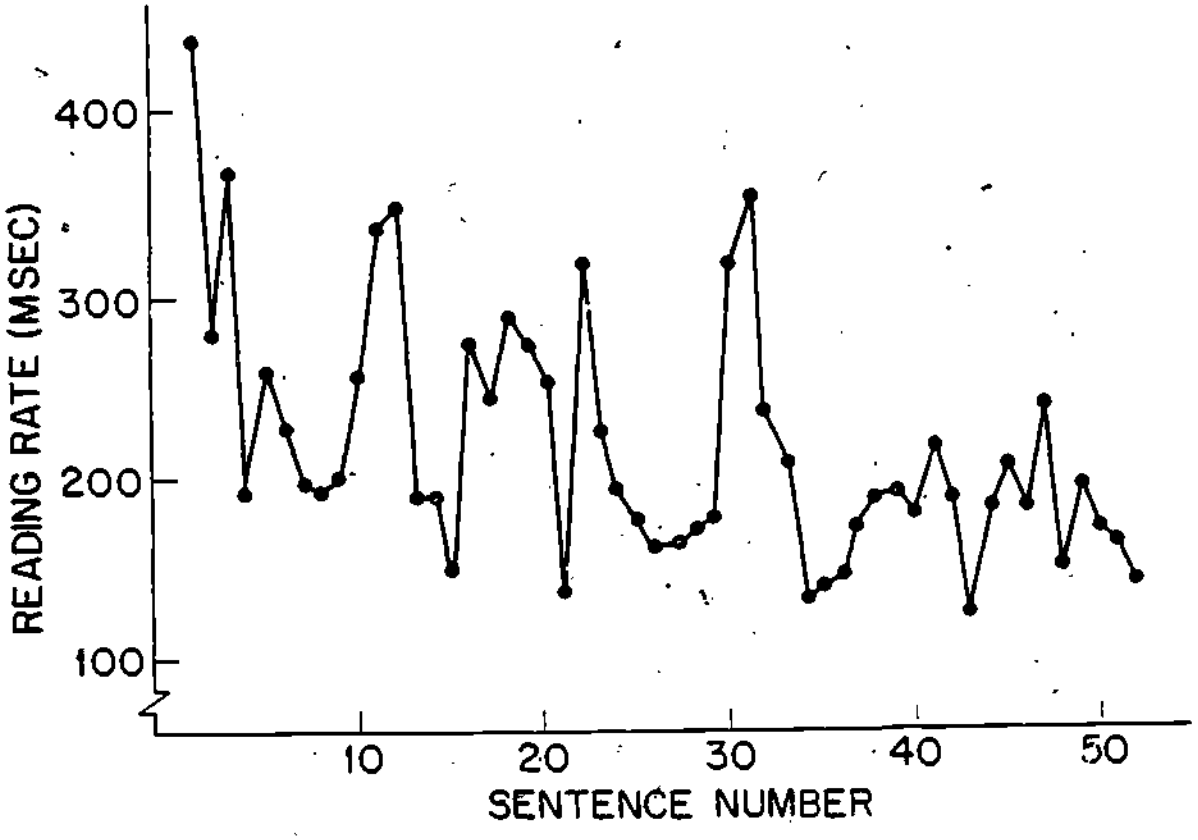
Figure Captions

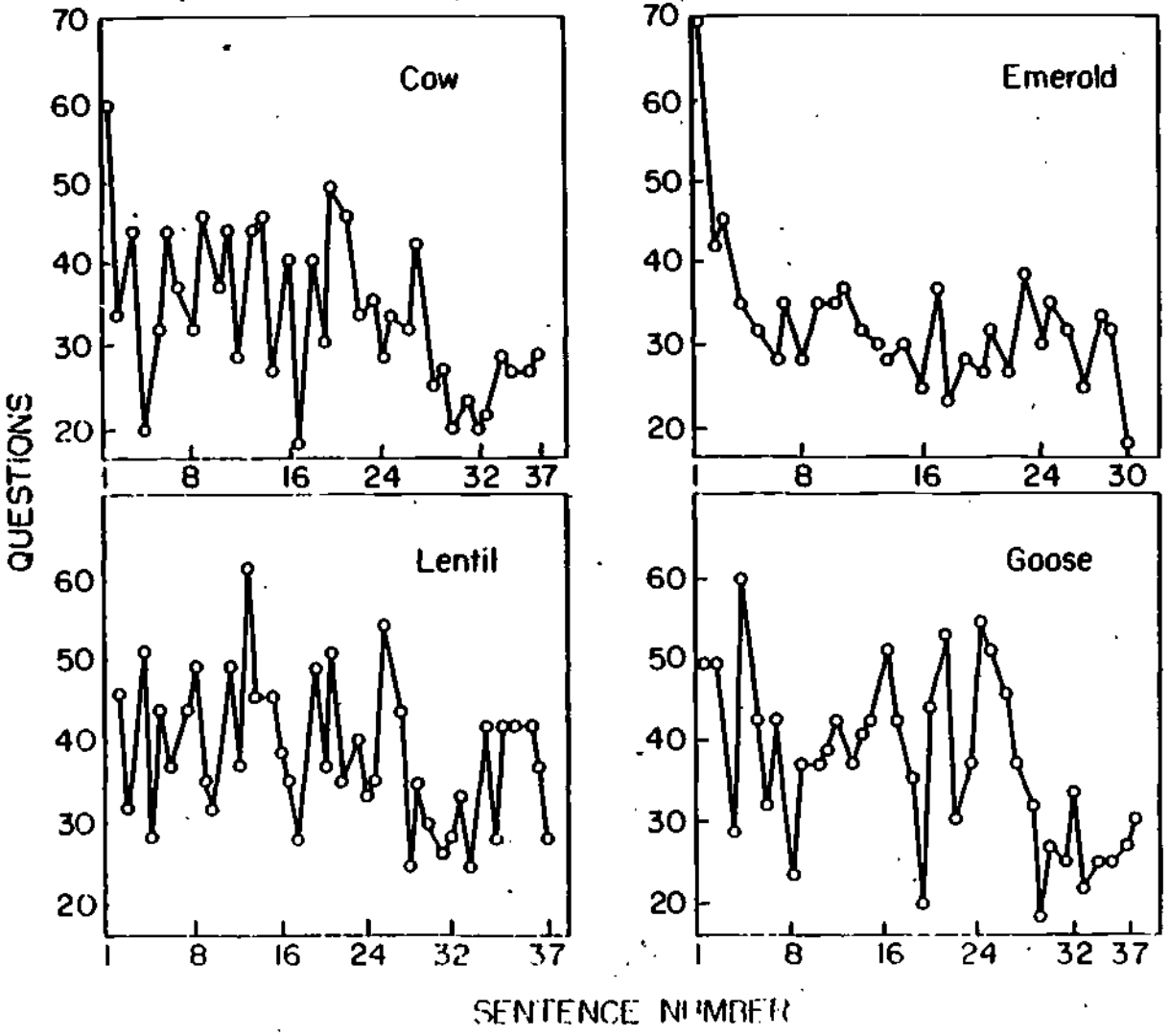
Figure 1. Conceptual schema for the reading process.

Figure 2. Mean reading rate for each sentence in Lentil

Figure 3. Total number of questions asked for each sentence in four simple stories.


Knowledge Base
Real World Knowledge
• Physical World
• Social World
Genre Knowledge
• Story Schemas
• Inductive Arguments
• Compare and Contrast Essays
Text Knowledge
• How Texts Are Assembled
Strategies
Text
Parsed Sentences
Comprehensive Processing
Working Memory
Current Lines of Thought
• Predicted Content and/or Structure
• Unanswered Questions or Confusians
• Criticisms
Knowledge Space
• Representation of Text Read So Far
Narrator Space


READING RATE (MSEC)
400
300
200
100
10
20
30
40
50
SENTENCE NUMBER


Cow
Emerold
Lentil
Goose
QUESTIONS
SENTENCE NUMBER
70
60
50
40
30
20
1
8
16
24
32
37
30

Question-asking as a Component of Text Comprehension

Gary M. Olson[1]
University of Michigan
Susan A. Duffy
University of Massachusetts
and
Robert L. Mack
IBM Watson Research Center

DRAFT: DO NOT CITE WITHOUT PERMISSION

To appear in: A. Graesser & J. Black (Eds.). The psychology of questions. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Question-asking as a Component of Text Comprehension

Gary M. Olson
University of Michigan
Susan A. Duffy
University of Massachusetts
and
Robert L. Mack
IBM Watson Research Center

In its primary mode of use, a question is a device for seeking new information that is to be related to an existing knowledge structure. When to ask a question, and exactly what to ask, are both symptomatic of the status of the knowledge structure at issue, as well as, no doubt, the general intelligence of the asker. We have all encountered the person (often ourselves!) who indicated they did not know enough about a topic to ask a question about it. Thus, intuitively, there is a link between one's knowledge or understanding of a topic and the ability to ask a question about it (e.g., see Miyake & Norman, 1979).

There is another connection between questions and comprehension. Educators and researchers have long suspected that approaching the comprehension of text with either general or specific questions in mind might facilitate understanding. There is a sizable research literature on this role of questions in understanding text (e.g., Anderson & Biddle, 1975; Frase, 1975). Questions of this type focus the reader's attention on exactly those pieces of information that are important to understanding

what the text is about. Since one of the problems faced by the reader is selecting the most relevant or important information from a text, appropriate questions can serve as a guide for this important process.

These two uses of questions in relation to understanding have an important relationship. Questions asked about a text are both an indication of having understood what has been read and a guide to the further understanding of what is about to be read. This suggested to us that questions asked by a reader while reading a text might be an especially informative kind of data for monitoring the reader's understanding of the text.

We have carried out a program of research aimed at finding what kinds of higher level cognitive processes readers engage in while reading simple texts. We felt one simple strategy for obtaining this kind of information would be to have readers think out loud while reading. We were motivated by a belief that intelligent reading has many affinities with problem-solving, a domain in which thinking-out-loud protocols have proved to be a useful research tool. A series of studies using this method have revealed a number of important phenomena about reading (Olson, Duffy & Mack, 1980, 1983; Olson, Mack & Duffy, 1981). One especially important finding has been that characteristics of the thinking-out-loud protocols correlate with silent reading time (see details in Olson et al., 1981, 1983), suggesting that the information obtained from this method is relevant to understanding the nature of text comprehension.

One of the things we noticed subjects doing while thinking out loud during reading was asking questions. The kinds of questions people asked and the places they asked them seemed to us indicative of important comprehension processes. This led us to conduct a specific study on the relationship between on-line question asking and comprehension. In this chapter we shall report a few highlights of this study. A more complete report of it will appear in Olson, Duffy, Eaton, Vincent and Mack (in preparation).

Let us summarize the general rationale for this study. The kinds of considerations we have sketched led us to believe that questions asked by subjects during the reading of a simple text would be diagnostic of important comprehension processes. It seemed plausible to assume that each sentence encountered in a text raises certain questions in a reader's mind and answers other questions raised by earlier sentences. We wanted to explore this supposition in more detail by collecting data on the kinds of questions readers ask following each sentence in simple stories.

This study used four tasks. The primary task was one in which readers asked questions after reading each sentence in the story. In another task a different group of subjects read the same stories silently while we timed their reading. These same subjects later recalled the stories. Finally, another group of subjects rated the importance of the constituents of the story. Four short simple stories (maximum length was 41 sentences) were

used as texts. They were all children's stories or simple folktales, and all were well-formed.

To better understand the results, a somewhat more detailed description of the four tasks is necessary:

1. Question-asking. All four stories were presented to 9 subjects. Each sentence in the story was typed on a card, and the subject worked his or her way through the deck of cards, asking questions that were raised in his or her mind as a result of having read that particular sentence. The subject was told to imagine that the story's author was present, and that the author was willing to answer any questions the reader had about the story at that point, except for the obvious question of what happens next. The subject was allowed to spend as much time on any sentence as he or she desired, but was asked not to reread any previous sentences or to look ahead. The questions were tape recorded and later transcribed. The number of questions asked for each sentence was tallied and pooled over subjects. In addition, the questions were classified in various ways.

2. Reading times. Sentence-by-sentence reading times were collected from 20 subjects. At the end of each story subjects wrote a brief (3 to 5 sentences) summary of the story.

3. Recall. The same 20 subjects were asked to recall the stories they had just read. They were presented with a brief descriptive title for each story, and were given unlimited time to try to recall as much as they could. They were asked to recall exact words, but were encouraged to guess if they could not remember exact words. Recall was scored by first doing a

propositional analysis of each story and then matching the subject's recall against this, using a gist criterion.

4. Importance. Seventeen subjects read each story and crossed out the 50% of the words, phrases, or sentences in the story they felt was least important. For each sentence in each story the proportion of words left in averaged over subjects provided a measure of the relative importance of that sentence.

It is useful to have a better picture of what the question-asking data look like. Table 1 shows typical questions for the first sentence of one of the stories. These questions are grouped into those asked by two or more subjects and those that are idiosyncratic to one subject. Of course, we were also interested in the sentence-by-sentence variation in the questions asked. Figure 1 shows the total number of questions asked for each sentence in each of the four stories. With the possible exception of EMERALD, there is noteworthy variation in the number of questions asked from sentence to sentence. In EMERALD, there were a large number of questions at the beginning and then a fairly flat distribution of questions thereafter. Keep this difference in mind, because EMERALD will not follow the pattern of other stories in some of our later analyses.

Insert Table 1 and Figure 1 about here

The first issue we addressed was whether the question-asking task is related to the reading times. We examined this by looking at the relationship between the total number of questions asked for each sentence in a story and the average reading time

for each sentence for those subjects who were reading silently. The expectation was that sentences which elicited a lot of questions would be especially salient during real-time processing, and therefore would be read more slowly by subjects who were reading silently. This hypothesis was confirmed. We conducted multiple regressions in which the average reading time per sentence was the dependent variable, and the predictor variables were sentence length, total number of questions, serial position, and importance. Only sentence length and number of questions emerged as significant predictors of reading time. In this analysis all four stories were entered, with story as a variable. There are two types of questions that occur: those that are asked by several subjects, and those that are idiosyncratic. We next asked whether these two types of questions contributed differentially to this outcome. The answer was no. A multiple regression with number of questions asked by two or more persons and idiosyncratic questions entered separately showed that both emerged as significant predictors. Table 2 shows the details of these analyses.

Insert Table 2 about here

When we carried out multiple regression analyses for each story individually, the results mirrored the overall analysis. In these regressions we included as predictors idiosyncratic questions and questions asked by two or more persons as well as total number of questions asked. For three of the four stories, at least one of these question counts emerged as a significant

predictor of reading time (in addition to number of syllables). The exception was EMERALD, for which the question data provided no significant predictor. As mentioned earlier, EMERALD was the story that showed little variation in number of questions asked across sentences.

So, number of questions asked accounts for a significant portion of the variance in sentence-by-sentence reading times. We next asked what relationship the question-asking task has with recall. And the answer was very simple: none. Table 3 shows the outcome of a multiple regression carried out on recall scores, and reveals that rated importance and serial position emerged as significant predictors of recall, while number of questions asked did not. This pattern is similar to other data which indicate that importance predicts recall (Meyer, 1975; Rintsch, 1974). Importance is not necessarily immediately perceived, but may result from having most or all of the final memory representation of the text. We conclude from this that the information being revealed by the question-asking task is more closely associated with the activities that occur during comprehension than with the form of the final memory representation constructed as a result of comprehension.

Insert Table 3 about here

This basic result confirms our initial supposition that the question-asking task would tap an aspect of what is going on in the skilled reader's mind while reading. The obvious question, of course, is what is it tapping? It is unlikely that a reader

who is reading silently is actually asking questions while reading. Rather, we believe that the question-asking task taps the kinds of informational needs a reader encounters while proceeding through a text. As each sentence is understood and added to a growing representation of the story, the reader revises and elaborates the set of information still needed to have the developing story make sense. These informational needs interact with what is presented in the next sentence to generate a new set of informational needs--or, if you will, a new set of questions--that guide the reader's comprehension through the succeeding parts of the text.

We have conducted a number of other analyses of these data that are discussed in Olson, Duffy, Eaton, Vincent, and Mack (in preparation). We have categorized the questions to see if certain types are more important than others. So far, the categories we have examined have not shown any differences. We have also looked to see whether or not questions asked are later answered by the story, and there are interesting relationships here. Many questions are in fact answered, though it varies somewhat by type. However, the number of questions answered by a particular sentence does not predict reading time or recall. We have looked at the information tapped in the question, and find that questions which are derived from new information contained in the current sentence are especially important in predicting reading times. These and other details of these data are interesting and important, and will be reported on fully in Olson et al. (in preparation).

The main findings of this study strongly suggest that the question-asking task is a useful indicator of processes which may be an important part of comprehension. The number of questions asked by subjects as they read through a story correlates with the amount of time spent on that sentence by other readers reading silently. Keep in mind that this result is with the obvious effect of sentence length removed. But number of questions does not correlate with recall. Thus, question-asking seems more closely related to the real-time processes that occur during reading than to the final product of comprehension that remains when reading is completed.

How general are these findings? We do not yet know. Clearly, we can only confine our conclusions to the reading of simple stories by reasonably sophisticated readers. Other types of stories, other types of texts, and other types of readers might yield quite different outcomes. But these initial results are promising enough to warrant the extension of this paradigm to these other situations.

References

Anderson, R.C., & Biddle, W.B. On asking people questions about what they are reading. In G. Bower (Ed.), The psychology of learning and motivation (Vol. 9). New York: Academic Press, 1975.

Frase, L.T. Prose processing. In G. Bower (Ed.), The psychology of learning and motivation (Vol. 9). New York: Academic Press, 1975.

Kintsch, W. The representation of meaning in memory. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Meyer, B. The organization of prose and its effect on memory. Amsterdam: North-Holland, 1975.

Miyake, N., & Norman, D. To ask a question, one must know enough to know what is not known. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 357-364.

Newell, A., & Simon, H.A. Human problem solving. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Olson, G.M., Duffy, S.A., Eaton, M.E., Vincent, P., & Mack, R.L. On-line question-asking as a component of story comprehension. Manuscript in preparation.

Olson, G.M., Duffy, S.A., & Mack, R.L. Applying knowledge of writing conventions to prose comprehension and composition. In W.E. McKeachie (Ed.), Learning, cognition, and college teaching. San Francisco: Jossey-Bass, 1980.

Olson, G.M., Duffy, S.A., & Mack, R.L. Thinking-out-loud as a method for studying real-time comprehension processes. In D.E. Kieras & M.A. Just (Eds.), New methods in reading

comprehension research. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1983.

Olson, G.M., Mack, R.L., & Duffy, S.A. Cognitive aspects of genre. Poetics, 1981, 10, 283-315.

Footnotes

'This research has been supported by a research grant from the National Institute of Education (NIE G-79-0133) and by a Research Career Development Award from NICHD (HD 00169) awarded to the first author. Communications about this work should be addressed to Gary M. Olson, Department of Psychology, University of Michigan, 330 Packard Road, Ann Arbor, Michigan 48104.

Table 1

Sample Questions from The Selling of the Cow

Questions asked by 2 or more subjects:	# Subjects
1) Who is Cromer?	2
2) What is Cromer like?	3
3) Did Cromer live alone?	5
4) When did this story take place?	5
5) Where was the farm?	4
6) Where was the hill?	4
7) Why was the farm on a hill?	2
8) How far up the hill was the farm?	2
9) How high was the hill?	3
10) What kind of farm was it?	4
11) What will happen to Cromer?	2
Idiosyncratic questions asked by only 1 subject:	
1) Does the fact that he lives on a farm have any significance?	
2) Does he farm for a living?	
3) Does he have another vocation?	
4) Is Cromer married?	
5) How old is Cromer?	
6) How far away were Cromer's nearest neighbors?	
7) Why did Cromer like to live on a farm?	

8) Are they going to roll something down the hill?

9) Did a lot of the dirt wash off the side of the hill so that Cromer couldn't have his crops?

10) What was Cromer's first name?

11) Was that Cromer's first name?

12) Then what was Cromer's last name?

13) Did Cromer have more than one name?

14) What kind of name is Cromer?

15) What does Cromer mean?

16) What nationality is Cromer?

Table 2

Multiple Regression Analyses of Reading Time in Question-Asking Experiment

	Regression Coefficient	Significance Level	Cumulative R^2
1. Predictors selected:			
Sentence length	130.22	.0001	.589
Total number of questions asked	26.44	.0001	.640
Predictors not selected:			
Serial position of sentence			
Importance			
2. Predictors selected:			
Sentence length	130.81	.0001	.589
Idiosyncratic questions	40.27	.0004	.626
Number of questions asked by two or more subjects	87.42	.0015	.652
Predictors not selected:			
Serial position of sentence			
Importance			

Note: Forward stepwise regression, dependent variable=mean reading time per sentence.

Table 3

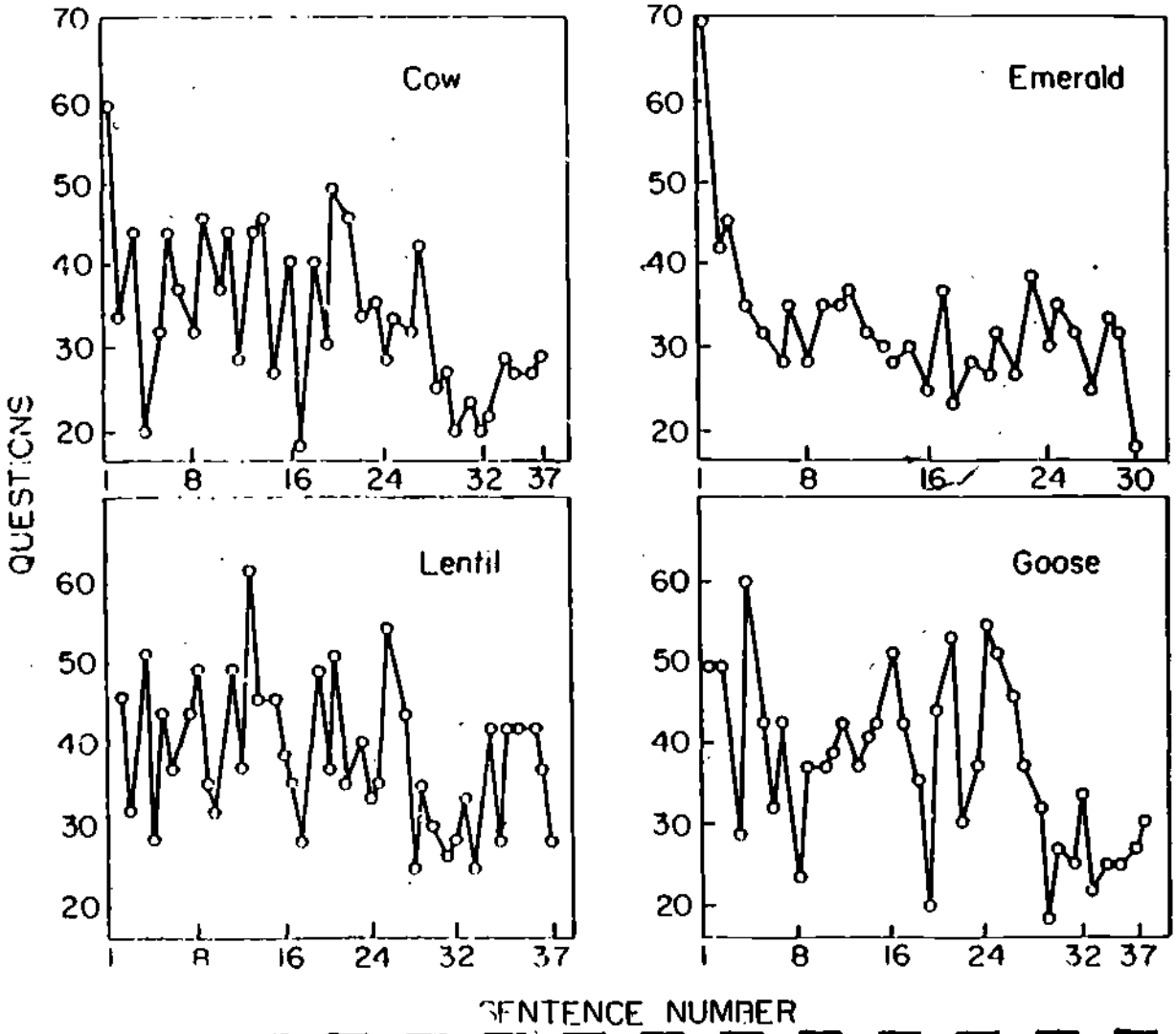
Multiple Regression Analyses of Recall in Question-Asking Experiment

	Regression Coefficient	Significance Level	Cumulative R^2
Predictors selected:			
Importance	341.12	.0001	.235
Serial position of sentence	-2.94	.0024	.263
Predictors not selected:			
Sentence length			
Total number of questions asked			

Note: Forward stepwise regressions, dependent variable=proportion propositions recalled per sentence.

Figure Captions

Figure 1. Total number of questions asked for each sentence in four simple stories.


Cow
Emerald
Lentil
Goose
QUESTIONS
SENTENCE NUMBER
70
60
50
40
30
20
1
8
16
24
32
37
30

The Role of Expectations in Sentence Integration

Susan A. Duffy

University of Massachusetts, Amherst

Annual Subject Index:

Comprehension

Text

Text

Processing

Sentence

Integration processes.

Abstract

In three experiments subjects read a series of narratives that varied in the degree to which they elicited an expectation that particular sentences would be followed up (High Expectation vs. Low Expectation texts). Subjects were faster to judge a target sentence to be unrelated when it followed a High Expectation text than when it followed a Low Expectation text. Subjects were slower to read an unimportant, expectation-violating sentence embedded in a High Expectation text. Results suggest that expectations are used to integrate upcoming sentences into the text representation. Correct expectations can help the integration process; incorrect expectations interfere. A third experiment found no evidence that these expectations took the form of highly specific predictions.

The idea that skilled readers regularly form expectations as they read is not a new one. Within the reading literature, claims have been made that reading is a "guessing game" in which readers make minimal use of the information on the printed page (Goodman, 1967; Haber, 1978; Smith, 1971). At least two kinds of expectations that a reader might use have been investigated with mixed results: expectations about individual words in a sentence (e.g., Ehrlich & Rayner, 1981; Eisenberg & Becker, 1982; McClelland & O'Regan, 1981), and expectations about upcoming syntactic categories (e.g., Fodor, Bever & Garrett, 1974; Mitchell & Green, 1978). The experiments reported here provide evidence for the use of expectations at a higher level of processing: expectations about upcoming events in a narrative. These expectations will be discussed in terms of the role they play in causal inferencing during reading.

The Problem of Causal Cohesion

If the reader's goal is to comprehend the sentences of a text as a coherent whole, the reader must find a way to integrate each successive sentence with the mental representation of the sentences already read. To integrate a sentence, the reader must find some way of linking the information in the sentence to a subset of the information presented earlier.

Two kinds of links are important in the current context.

The first are the referential links established between an anaphoric phrase and its antecedent. The second are links which establish what Keenan (Note 1) refers to as causal cohesion among propositions. An example can make this distinction clear.

1a. Johnny had blonde hair.

1b. He bought a harmonica.

2a. Johnny wanted to make music.

2b. He bought a harmonica.

In both of these sentence pairs, a referential link is established between "He" in sentence b and "Johnny" in sentence a. The two sentences in the first pair, however, seem to be unrelated facts about Johnny. In contrast, the second pair are causally related in the sense that the want expressed in 2a is the "cause" for the action taken in 2b. To fully understand these two sentences the reader must establish both the referential and the causal link. If the reader only establishes the referential link in the second pair, an important aspect of the intended meaning will be lost.

The discovery of causal cohesion is especially important in narratives. When a reader makes sense of a narrative text, a major goal is to build a representation of the causal links among events (what Wilensky, Note 2, terms "explanation-driven understanding"). The reader wants to be able to explain why certain events occurred (in terms of causes and enabling conditions) and what happened as a result (consequences). Thus

the reader might be expected to pay attention to causally important information. This is information which needs explanation or which might provide an explanation for upcoming events in the narrative. This claim is given some support by research of Cirilo and Foss (1980) which showed that readers spend more time reading important sentences in stories, sentences which make up the causal chain. Important sentences, especially early in a story, also happen to be sentences which will have consequences later on. This means these sentences will be crucial for establishing causal cohesion later on (when the sentences conveying the consequences are encountered).

Current models of sentence integration focus on the search for referential links as the key to the process of relating the current sentence to what has been read (Clark & Haviland, 1977; Clark & Sengul, 1979; Garrod & Sanford, 1977; Kintsch & van Dijk, 1978; Lesgold, Roth & Curtis, 1979). Yet as the above account makes clear, the discovery of causal cohesion is also crucial to comprehension.

One might imagine that establishing referential cohesion is a major part of finding causal cohesion. For example, the search for antecedents for definite noun phrases in the current sentence might lead directly to the information in the earlier text needed to establish causal cohesion. In fact, however, causal cohesion can exist between sentences with no explicit argument overlap. An example can make the problem clear.

John was eating in the dining car of a train.
The waiter brought him a large bowl of bean soup.
Suddenly the train screeched to a stop.
The soup spilled in John's lap.

How does the reader comprehend the final sentence of this narrative? Under current models of sentence integration, the noun phrases "The soup" and "John's lap" will be used as search cues to find earlier, related information. Both of these cues will guide the search back to the second sentence, where reference is made to John and to the soup. As a result, according to current models, the last sentence will be integrated with the second sentence.

This procedure, however, misses the crucial causal link between the third and fourth sentences. The antecedent search process does not help in finding this causal link because of the lack of argument overlap between the third and fourth sentences in the text. The question, then, remains of how the reader does decide to consider generating a link between the third and fourth sentences in the absence of argument overlap.

A complete answer to this question must await a theory of causal inferencing. It is possible, however, to speculate about mechanisms that might be useful in guiding causal inference. One possibility is that the reader is using expectations to guide the process of finding causal links. If the reader has an expectation that a particular sentence contains a causally important event which will be immediately followed up by the

writer, then that sentence is a likely locus for integrating the next sentence, regardless of where else the antecedent search may lead. Consider the third sentence of the train text, "Suddenly the train screeched to a stop." This sentence seems likely to play a causal role in the events to come; it is the kind of event which is likely to have causes and/or consequences which are important to the narrative. It is also the kind of event that should be immediately followed up by the writer if the narrative is to be well-formed. Thus it is reasonable to hypothesize that the reader's response to such a sentence will be to form an expectation that upcoming sentences will relate to it.

Eliciting an Expectation

What kinds of events elicit expectations that they will be followed up? The likely candidates are events which violate a currently active script (Schank & Abelson, 1977). For example, trains do not typically screech suddenly to a stop. The event violates the "riding a train" script and it does not fit the "restaurant" script that may also be activated. As a result, the reader can infer that this event may be causally important. It is also likely that certain kinds of events access causal connections in long-term memory. For example, the reader probably has stored in memory information about the general concept of a sudden change in the velocity of a moving object. This general concept has strong links to likely causes (e.g., hitting a barrier, an agent applying brakes, etc.) and

consequences, e.g., (the displacement of supported objects). As soon as the train event is recognized as an instance of the more general concept, the reader has access to causes and consequences that suggest the event could be causally important (see Fahlman, 1979, for some ideas about modelling hierarchies of actions with cause and consequence links).

In contrast suppose the key sentence in the train text is changed so the whole text now reads:

> John was eating in the dining car of a train.
> The waiter brought him a large bowl of bean soup.
> The train slowed entering a station.
> The soup spilled in John's lap.

The text no longer contains a causally important event prior to the last sentence. The slowdown of the train is a normal event accounted for within the train script. Furthermore it does not seem likely to access a long term memory concept node which is bristling with cause and consequence links. It should be clear, then, that the expectations discussed here are generated selectively, not for every sentence of every text.

In short, a general expectation that more will be said about a particular set of propositions is elicited when those propositions are perceived to be of immediate causal importance. In its most general form, this type of expectation consists of the tagged text propositions which need to be followed up immediately. These tagged propositions may be linked to a number

of possible causes and consequences activated in long term memory. The set of tagged text propositions plus the activated concepts will be referred to as a general expectation. It is general in the sense that a number of possible next events will fulfill the expectation.

In addition to this general expectation, the reader may be able to form a very specific prediction about the next event in the narrative. For example, for the train text the reader may be able to predict that the soup will spill. Making this prediction involves the specific instantiation of one of the general causes and consequences activated for the event, the consequence of object displacement. Thus, the generation of a specific prediction may be viewed as a two-step process in which a general expectation is generated first, and then one aspect of that expectation is further specified.

It should be pointed out that the reader will not always be able to generate a specific prediction. For example, suppose the key sentence in the train text is changed to read, "Suddenly there was a high-pitched electronic tone." This is certainly an event likely to be followed up in the text. The writer must say something about its causes or consequences. Thus the reader is likely to form a general expectation as described earlier, but is unlikely to generate a specific prediction.

The Role of Expectations in Sentence Integration

The model to be developed here assumes that readers regularly form expectations (general and/or specific) as they

read narratives. Once formed, an expectation is used in the attempt to integrate upcoming sentences. Specifically, the content of the expectation is given priority in the reader's search for a causal link between the next sentence and the already-read text. A prediction follows from this account: when the reader has generated an expectation that should be immediately fulfilled, an attempt will be made to create a causal link between the upcoming sentence and the expectation. This attempt will be made even in the absence of argument overlap. For example, upon reading that the train suddenly stopped, the reader will attempt to find a causal relation between the upcoming sentence and this event. If the reader has formed a specific prediction, this process may involve an attempt at relating the sentence to the prediction. If the reader only has a general expectation, the process may involve constructing a relation between the sentence and the text propositions tagged as part of the expectation.

The expectation identifies that subset of text information which should be relevant to integrating the upcoming sentence. If the expectation is correct, then a causal link can easily be created, and the reader can bypass some of the search and inferencing that might otherwise be required. If the expectation is incorrect, then the link between the upcoming sentence and the text will not be so easily found because the reader will waste time trying to find a causal link where none exists.

Three experiments are reported which test the major

hypotheses about the role of expectations in sentence integration. The first two experiments test a general claim of the model without distinguishing between general and specific expectations. This claim is that an expectation provides a focus for current causal inferencing and as a result is given priority in the attempt to integrate upcoming sentences with the prior text. Experiment 1 tests the hypothesis that an expectation helps in the attempt to integrate when it is correct. Experiment 2 tests the hypothesis that an expectation interferes when incorrect. Finally, Experiment 3 turns to the general vs. specific distinction. It examines the content of the expectation itself, testing the hypothesis that readers are actually forming specific predictions (e.g., the soup will spill) whenever possible as opposed to more general kinds of expectations.

EXPERIMENT 1

In this experiment subjects read a series of short text fragments. At the end of each text, they were asked to judge whether a target sentence was a possible "next sentence" for the text. Response time to the target sentence was the dependent variable. Texts were designed either to elicit or not to elicit an expectation at the sentence immediately preceding the target sentence. Response time to the target sentence was expected to reflect processing in the presence or absence of an expectation.

A set of text fragments similar to the train text was

created. Each text was built around an everyday activity for which the reader might have a script (e.g., riding on a train, fixing food in a restaurant, hiking in the forest). Each text had a High Expectation (HiE) version designed to elicit a strong expectation that certain sentences would be followed up. Expectations were created by including a key causal event (e.g., the train screeching to a stop) or by including a causal trait (e.g., anger) accompanied by appropriate enabling conditions (e.g., using a knife in the kitchen). The key causal sentences departed from the script and will be referred to as the expectation sentences. To create the Low Expectation (LoE) version of each text the expectation sentences were modified so that they fit in with the normal, script events of the rest of the text.

For each text two target sentences were written. One sentence followed directly from the expectation sentences in the HiE version of the text. The other sentence was designed to be unrelated to the text. Examples are given in Table 1. The judgment task was intended to be a simple one for the subjects. As a result Unrelated targets were chosen to be obviously not related to any propositions mentioned earlier in the text.

Insert Table 1 about here

If expectations can provide a focus for causal inferencing, then response time to the target sentences for the HiE texts

should be faster than for the LoE in both Related and Unrelated conditions. The reasoning behind this prediction in the Related condition should be clear. For the HiE texts subjects have focused on certain propositions which are exactly those needed for the causal inference allowing integration of the Related sentence. For the LoE texts subjects have no such focus, and therefore they must search through the text to find appropriate information to allow integration.

The prediction for the Unrelated condition follows the same sort of reasoning. In the HiE condition the reader attempts to relate the sentence to the expectation. Upon finding no relation the reader can safely judge the sentence as Unrelated. This allows the reader to respond without taking the time to search the rest of the text for possible relations. In the LoE condition the reader does not have such a focus and thus must search more of the text before determining that the sentence is indeed unrelated.

The Unrelated condition is crucial for ruling out two important alternative hypotheses which could fully account for the predicted pattern in the HiE-Related condition. The first hypothesis is a backward inference hypothesis. Suppose the reader does not generate any expectations at all for a text. In order to integrate the Related sentence, some inferencing must be carried out to determine the causal link between prior text information and the sentence. The causal antecedents are much more likely as causes in the HiE texts (e.g., a train screeching

to a stop is a much more obvious cause of soup spilling than is a train slowing to enter a station). As a result, it is reasonable to expect that the inferences necessary to determine causality (and hence to find a relation) will be more quickly made in the HiE condition for the Related sentences.

This alternative hypothesis, however, would have trouble accounting for a difference in HiE vs. LoE response times in the Unrelated condition. The target sentence in this condition bears no causal (or referential) relation to any sentence in either the HiE or LoE versions of a text. Thus a backward inference hypothesis which claims that no prior information is given special focus would predict no difference in response time. If a difference is found, it will lend support to the claim that at least some of the effect in the Related condition is due to the presence of an expectation.

A second hypothesis will also be ruled out given an effect in the Unrelated condition. This hypothesis claims that an expectation is elicited by every text. This hypothesis is reasonable if it is assumed that expectations are simply concepts activated automatically for every sentence. A predict-at-every-sentence hypothesis could account for a HiE vs. LoE difference in the Related condition in the following way. For the HiE version of a text the expectations generated are likely to be relevant to integrating the target sentence; for the LoE version the expectations are unlikely to be relevant (e.g., for the LoE train text expectations might center around passengers getting on and

off at the station). The hypothesis cannot predict an effect in the Unrelated condition because the expectations for both the HIE an LoE textn would be equally irrelevant to attempts at integrating the Unrelated target sentence.

Methods

Subjects. Twenty-four subjects from the University of Michigan community participated. Each was paid $3.50 for a 30 to 40 minute session.

Materials. The texts were 40 short narratives. Each text was incomplete and could be considered the beginning of a longer story. Each text had two versions as shown in Table 1. In one version the reader was given one or two critical pieces of information intended to elicit an expectation at the end of the text (the expectation sentences). This version is the HIE version of the text. In the other version of each text, the expectation sentences were modified to be less likely to elicit an expectation at the end of the text. This second version is the LoE version of the text.

For each text a Related target sentence was written. These sentences were eight to ten syllables long, with a mean length of 9.1 syllables. The sentence always directly followed from the expectation sentences in the HIE version of the text. Argument overlap between the Related sentence and both versions of a text was always the same.

All texts were then paired such that the Related target sentence for one member of the pair could be used as the

Unrelated target sentence for the other member of the pair. This pairing was not random; care was taken to insure that there was no argument overlap between text and Unrelated sentence and that the Unrelated sentence for a text was as convincingly unrelated as possible. The result of the pairing was a set of stimulus materials in which each text and each sentence appeared in all possible conditions.

Design. There were four experimental conditions formed by the crossing of text version (HIE vs. LoE) with target type (Related vs. Unrelated). A given subject saw ten texts in each of the four conditions. Across the full experiment each text and each target sentence appeared equally often in all four conditions, although for a given subject each text and target sentence appeared in only one condition.

Procedure. The experiment was controlled by a Digital PDP 11/34 computer. Subjects were run individually, in soundproof booths. The texts were displayed on a Hewlett-Packard 2621A Interactive terminal connected to the computer. Subjects responded by pressing keys on a keyboard in front of them.

Each subject read all 40 texts plus an additional ten practice texts which were placed at the beginning of the experiment.

Subjects began the experiment by pressing the space bar on the keyboard. The first text was displayed in full on the CRT. Subjects had unlimited time to read the text. When they had finished reading, subjects again pressed the space bar. The text

was replaced by a single target sentence. Subjects read the sentence and decided as quickly as possible whether it was a reasonable next sentence, using the "/" and "Z" keys to indicate their response. After subjects responded, a feedback message appeared on the CRT for 1000 msec. The feedback message was automatically replaced by a message telling the subject to press the space bar for the next text.

Subjects were instructed not to deliberate over the target sentences, but rather to respond as quickly as possible but not so fast that they made lots of bad judgments. They were told that they should respond "Yes" if the sentence was a possible next sentence, "No" if it was not. It was emphasized that subjects should not worry about whether it was the best possible next sentence.

Because the criterion for deciding that a sentence was Related or Unrelated might be expected to vary from subject to subject, feedback was given as guidance for setting the criterion. Relatedness judgments about the target sentences had been informally collected from colleagues. When a subject's response agreed with the judgment of this earlier group, the word AGREE appeared on the screen. If a response disagreed, DISAGREE appeared.

Results

For each subject a mean response time to the target sentence was calculated for each of the four conditions, and for each text

a mean response time was calculated for each of the four conditions. These means were used as the observations in the analyses. Trials in which subjects disagreed with the expected judgment were excluded from the calculation of these means. In addition trials more than 2.5 standard deviations from a subject's mean for a given response type ("Yes" and "No") were excluded as outliers (1.9% of the data). Tests based on subject variability will be referred to as F_1; those based on item variability will be referred to as F_2. Planned comparisons are based on subject variability, the Bonferroni t procedure is used with the critical significance level adjusted according to number of comparisons made.

Mean response times and disagreement rates for each condition are displayed in Table 2. In the analysis of response times, the effect of text version was significant; subjects were faster to respond to the target sentence when it followed a HiE text than when it followed a LoE text ($F_1(1,23) = 60.46$, $p < .0001$, MSe = 93735; $F_2(1,39) = 35.04$, $p < .0001$, MSe = 93470). In addition the interaction of text version and target sentence type was significant ($F_1(1,23) = 20.60$, $p < .0004$, MSe = 31632; $F_2(1,39) = 8.74$, $p < .006$, MSe = 154996). Planned comparisons revealed that the mean response time to the HiE-Related target sentences was significantly faster than the mean response time to the LoE-Related target sentences ($t(23) = 6.74$, $p < .0001$). The mean response time to the HiE-Unrelated target sentences was faster than the mean response time to the LoE-

Unrelated target sentences ($t(23) = 4.32$, $p < .0005$).

Insert Table 2 about here

In the analysis of disagreement rates, the effect of text version was significant. The probability of disagreement was higher for the target sentences following the HiE texts than for sentences following the LoE texts ($F_1(1,23) = 8.95$, $p < .007$, MSe = .561; $F_2(1,39) = 7.89$, $p < .008$, MSe = 107). The interaction of text version and target sentence type was significant ($F_1(1,23) = 10.99$, $p < .004$, MSe = 97; $F_2(1,39) = 14.06$, $p < .001$, MSe = 126). The disagreement rate for the LoE-Related condition was significantly higher than the LoE-Unrelated condition ($t(23) = 2.87$, $p < .01$) and higher than the HiE-Related ($t(23) = 4.13$, $p < .001$). The HiE-Unrelated and LoE-Unrelated did not differ significantly; nor did the HiE-Related and the HiE-Unrelated.

Reading times for the texts themselves were compared for the HiE and the LoE conditions. Two kinds of analyses were conducted. First, for each subject the mean reading time for the HiE and for the LoE conditions was calculated. A paired t-test showed no significant differences in these means ($t(23) = .05$). The text versions do differ, however, in mean length: the HiE versions have a mean of 46.85 syllables; the LoE a mean of 43.02 syllables. As a result a second analysis was conducted which took syllable length into account. For each subject two regressions were carried out, one for the HiE texts, one for the

LoE. The dependent variable was reading time for the text version; the independent variable was the number of syllables in the text version. Analyses of the resulting intercepts and slopes also revealed no significant differences for the HiE vs. LoE conditions (intercept analysis: $t(23) = -.05$; slope analysis: $t(23) = -.31$).

Discussion

Subjects were faster to respond to both the Related and the Unrelated target sentences when they followed a HiE text. This pattern of results is consistent with the general hypothesis introduced earlier. A correct expectation provides a focus for causal inferencing, presumably allowing the subject to bypass some of the processing normally required in the attempt to link a sentence to those preceding it. The results also support an assumption made about this type of expectation: such expectations are generated selectively, and not for every sentence of every text.

As argued earlier, the Unrelated condition is important for providing evidence against two plausible alternative hypotheses: the backward causal inference hypothesis and the predict-at-every-sentence hypothesis. While both hypotheses can predict the pattern of results found in the Related conditions, these two hypotheses cannot predict the difference found in the Unrelated condition.

Because the Unrelated condition is a critical one, it is worth considering what relation it has to situations typically

encountered during "normal" reading. It seems clear that skilled readers frequently encounter sentences in stories which vary in how strongly they are predicted by the previous text (e.g., HIE-Related vs. LoE-Related). Good writers, however, do not deliberately include Unrelated sentences in their texts. Nevertheless it is likely that readers have experienced the phenomenon of encountering a seemingly unrelated sentence. Frequently a sentence is read with less attention than it deserves, and a word is misidentified or a phrase incorrectly parsed. Sometimes a sentence is actually ambiguous, and the wrong interpretation is chosen by the reader. Such ambiguous texts have been studied by Collins, Brown and Larkin (1980) and by Rumelhart (Note 3). As a result of misinterpretation of the early ambiguities in such texts, a later sentence will seem incongruous or unrelated.

Skilled readers must be sensitive to such incongruity because it is a good indicator that the comprehension process has failed and remedial action is necessary. For example, it may be necessary to reread earlier sentences to find where the misreading occurred, or it may be necessary to reinterpret an earlier sentence which is in fact ambiguous. Skilled readers must have procedures for detecting and correcting such instances of miscomprehension. In fact detailed protocols of such corrections were collected by Collins et al (1980) and by Rumelhart (Note 3). It seems, then, that the task used here can be expected to tap those procedures readers normally use in

reading. One conclusion to be drawn from the results is that an expectation can help a reader catch comprehension problems.

There is, however, a major difference between the reader's response to an Unrelated target sentence in the HIE condition and to an incongruous sentence encountered in normal reading. In this experiment the Unrelated sentences were not supposed to fit in, and the subject knew that. As a result, when the subject encountered a target sentence that did not relate to the expectation, it was a good bet that the sentence must be Unrelated. On this basis, the subject could make a "No" response, making no further attempt at integrating the sentence. In normal reading, of course, readers would take the time to go back to figure out where the misreading occurred and how the incongruous sentence actually fit in.

One notable characteristic of the results is the fact that the difference between the mean response time for the HIE and LoE conditions is about 3 1/2 times larger in the Related than in the Unrelated condition. Two complementary accounts of this interaction can be given. First, it is likely that the HIE response times consist of a mix of at least two different kinds of trials: trials on which the appropriate expectation is formed and trials on which no expectation is formed. In the Related condition a faster response is expected for both types of trials, given the predictive and backward inferencing processes described earlier. In the Unrelated condition, however, the latter type of trial (where no expectation is formed) will not result in a

faster response. A faster response will occur only on those HIE trials where an expectation is formed. Hence the proportion of fast trials is lower in the Unrelated condition.

A second possible reason for the interaction in response times is suggested by the analysis of disagreement rates. It seems to be especially difficult to decide on a response for a Related target sentence when it follows a LoE text. In this condition subjects will discover argument overlap between the target sentence and the text; this suggests that a relation is possible. The causal links required, however, are much less likely than in the HIE-Related cell. As a result, the decision stage as well as the search and inference stage may be lengthened. For example, subjects may spend extra time deciding whether the causal relation they have inferred between the train slowing down and the soup spilling is reasonable enough to allow a Related response.

One additional hypothesis might be proposed to account for the overall pattern of results in this experiment. This is a "general arousal" hypothesis. Under this hypothesis, the HIE texts do not elicit specific expectations. Rather, they prime subjects to respond faster to whatever stimulus next appears. This hypothesis is difficult to specify, but it is based on the observation that in this experiment subjects are always faster to the HIE targets no matter what relation exists between target and text. One could claim that the HIE texts somehow leave subjects in a "higher state of arousal," and thus response time is always

faster. This alternative hypothesis is ruled out by the results of Experiment 2.

EXPERIMENT 2

The results of Experiment 1 provide support for the claim that expectations are formed during reading, that they are formed selectively and not for every text, and that they can be used in the attempt to relate the upcoming sentence to the previous text. In Experiment 1, subjects were reading with the assumption that they would encounter unrelated sentences. As a result, they could make use of their expectations to quickly reject a target sentence which bore no obvious relation to the expectation. Suppose, however, subjects were reading with the assumption that all sentences could be integrated into the text representation (the assumption made for normal reading). How might the presence of an expectation affect integration processes for a sentence which was related to the text but which did not fulfill the expectation? Experiment 2 was designed to address this question.

If expectations always become involved in the attempt to integrate the upcoming sentence, then time may be wasted in processing a sentence which does not relate to the expectation. Thus an incorrect expectation might be expected to interfere with sentence integration. This hypothesis will be referred to as the interference hypothesis.

For example, consider the following modification in the last

sentences in the train text:

Suddenly the train screeched to a stop.

The waiter offered John some coffee.

How is the last sentence processed? According to the interference hypothesis, an attempt will be made to figure out how the waiter offering coffee relates to the train screeching to a stop. This commitment to try to relate incoming sentences to the expectation is costly (in processing time) when there is no relation intended by the writer. Thus this hypothesis predicts slower integration times for sentences which violate an expectation.

An alternative to the interference hypothesis might claim that expectations are available to be used if relevant, but they do not become involved in time-consuming processing if irrelevant or incorrect. Under this view, expectations might consist of concepts automatically activated in long term memory. The script activation model of Bower, Black and Turner (1979) is a model of this type. In this type of model a script event encountered in a text is assumed to activate upcoming events in the long term memory representation of the script. These activated events, or expectations, can help in processing the next text event, but if the next event is not a script event, the activated script events do not interfere with sentence integration.

Consider how this hypothesis might account for processing in the train text. When the final sentence is the event of soup spilling, the concept "spill" is directly related to one of the

concepts likely to be activated as part of the expectation (e.g., "displacement" is one of several causes and consequences activated as part of the general expectation). Thus the expectation will become involved in processing through activated long term memory pathways. In contrast, when the final sentence is the event of the waiter offering coffee, the expectation will not be accessed because the concepts in this sentence have no direct relation through long term memory links to the expectation.

A second type of non-interference hypothesis could also be developed. Most narratives convey some script information as well as major events which form the causal chain. The reader may first check whether the current sentence relates to the active script and if it does not, only then go on to check its role in the causal chain (i.e., its relation to the expectation). This account would also predict that the expectation would not become involved in processing the sentence in which the waiter offers coffee.

Experiment 2 was designed to distinguish between the interference and the non-interference hypotheses. Texts similar to those used in Experiment 1 were constructed. As in Experiment 1, each text had a script activity as a theme. The HIE version of each text contained sentences designed to elicit an expectation that they would be followed up. In the LoE version these expectation sentences were modified so that they did not elicit strong expectations that they would be followed up. For

the HiE version, two pairs of target sentences were written: an expected (Exp) pair and an unexpected (Unex) pair. Examples are given in Table 3. The expected target sentence pair followed from the expectation sentences in the HiE version (soup spilling in John's lap); the unexpected target pair was a normal event in the script (the waiter offering John coffee). The same target sentehce pairs were used for the LoE version. The labels "Expected" and "Unexpected" will be used throughout, although it should be clear that neither label actually applies to the target sentences when they follow the LoE versions of the texts.

Insert Table 3 about here

Both the interference and non-interference hypotheses predict faster reading times for the HiE-Exp target sentences than for the LoE-Exp. Only the interference hypothesis, however, predicts that processing of the Unex target sentence will be slowed when an expectation is present (i.e., reading time for the target sentence in the HiE-Unex condition will be slower than in the LoE-Unex). The non-interference hypotheses predict no differences in integration time for the target sentence in the HiE-Unex vs. LoE-Unex conditions.

It should be noted that a backward inference model will also predict no differences in the Unex conditions. Consider again the train text example. For both the Unex conditions antecedent search processes will lead back to the point where the waiter

served John earlier. The target sentence is easily integrated with this earlier information because it fits reasonably with the restaurant script. In neither condition does the target sentence relate directly to any later sentences. Thus a backward inference model will predict no differences in integration time for the HiE-Unex vs. LoE-Unex conditions.

Methods

Subjects. Sixty-five subjects from the University of Michigan community participated in this experiment. Of this number, 49 participated in a rating task to validate the texts; 16 participated in the reading time experiment. Subjects were paid $3.50 for an hour's participation.

Materials. Twenty-four of the texts from Experiment 1 were revised to meet the requirements of both Experiments 2 and 3. Each text was rewritten so that the HiE version of the text elicited much stronger expectations than did the LoE version. Furthermore, for each text a specific prediction could be identified. This prediction involved a non-human argument mentioned earlier in the text (e.g., soup), and was not highly semantically related to the information unique to the HiE version of the text (e.g., in isolation, a train screeching to a stop is not semantically related to soup spilling).

In order to rewrite and validate the texts, a preliminary rating task was run. Subjects were run in groups of two to six. Each subject was given a booklet containing one version of each of the 24 texts. Subjects were asked to write a sentence

conveying what was most likely to happen next and to rate how likely they thought their prediction was. A five-point scale was used. A rating of one indicated a "strong" prediction ("The reader has been set up to expect this event to occur."). A rating of five indicated a "weak" prediction ("The reader cannot tell what might happen next."). Subjects were instructed to guess if they did not feel they could predict what would happen, and to use the rating to indicate their lack of certainty.

Initially, the results from the rating task were used as a guide for rewriting the texts. The data collection and rewriting were carried out iteratively. Some texts required no rewriting; some required one or two rewritings. As a result, the number of subjects actually responding to the final versions of each text varied. But no text had fewer than six subjects responding per final version.

For Experiment 2, the analysis of the ratings is central (other analyses will be discussed for Experiment 3). The mean rating for the HiE versions of the texts was 1.67; the mean rating for the LoE versions was 2.92. The ratings differed significantly ($t(23) = -10.1$, $p < .0001$), confirming that the HiE texts do elicit stronger expectations about upcoming events.

For the final versions of each text, two sentences were written as the target sentences (see Table 3). The Expected target sentences (Exp) followed from the expectation sentences in the HiE versions of the text; the Unexpected target sentences (Unex) did not. The target sentence was always 11 syllables

long. An additional sentence of eight syllables (the secondary target) was written to follow the target sentence in case processing tended to spill over to the next sentence.

The Exp target sentences always conveyed important events in the story. The Unex target sentences conveyed unimportant events which related to the information contained in the portion of the texts which the HiE and LoE versions had in common. That is, the antecedents for the Unex sentence always appeared in identical sentences in the HiE and LoE versions of a text; the links joining the Unex target sentences to the text were intended to be the same in the two versions.

Design. Each subject read one version of each of the 24 texts. There were four experimental conditions formed by the crossing of text version (HiE vs. LoE) with target type (Exp vs. Unex). A given subject saw six texts in each of the four conditions. Across the full experiment each text appeared equally often in all four conditions. A different random order of presentation of texts was used for every four subjects.

Procedure. The experimental equipment was the same as in Experiment 1. Texts were displayed one sentence at a time on the CRT. Subjects controlled the presentation of each sentence by pressing the space bar on the keyboard in front of them. Each press of the space bar caused the current sentence to be erased and the next sentence to be displayed. Subjects read through the whole text including the target sentences; the target sentences were not identified in any way. After subjects had read the

secondary target sentence, a test sentence was presented, signaled by a large arrow which appeared above it. Subjects indicated whether the sentence was true or false of the text they had just read, responding by using the "z" and "/" keys. The test sentence was included to make sure subjects were actually reading for comprehension. The test sentence was always a shortened version of the expected event for the HiE version of the text. Feedback was given after the response.

Subjects read a total of 56 texts. Twenty-four were the experimental texts. Eight texts at the beginning were practice texts. Twenty-four filler texts were included to dilute the impression that highly unexpected events happened frequently.

Results

Means were computed as in Experiment 1. Mean reading times for the target sentence for each condition are given in Table 4. The means were submitted to two overall ANOVAs. In the ANOVA by subject, subjects were nested within group (defined by four different assignments of texts to conditions); both factors were crossed with text version (HiE vs. LoE) and sentence type (Exp vs. Unex). In the ANOVA by texts, sentences were nested within both text group and sentence type; all three factors were crossed with text version. In both analyses the text version by sentence type interaction was highly significant ($F_1(1,15) = 13.21$, $p < .005$, $MSe = 71032$; $F_2(1,40) = 25.61$, $p < .001$, $MSe = 51438$). In the text analysis the interaction of text group and text version was significant ($F_2(3,40) = 4.21$, $p < .05$, $MSe = 51438$).

No other effects were significant.

Insert Table 4 about here

Planned comparisons were used to examine the interaction of text version and sentence type. The Bonferroni t procedure was used, with a critical significance level of .0125 for each individual comparison. For the HiE versions, the Exp sentences were faster than the Unex ($t(15) = 3.26$, $p < .01$). For the LoE versions, the difference in reading time for the Exp vs. Unex sentences was marginal at best ($t(15) = 2.63$, $p < .025$). For the Expected target sentences reading time was faster following the HiE text versions ($t(15) = 4.06$, $p < .005$). Finally, for the Unexpected target sentences the reading time was faster following the LoE text versions ($t(15) = 3.11$, $p < .01$).

The mean reading times for the secondary target are also given in Table 4. The means for the secondary target sentence have the same interaction pattern as those for the target sentence. The differences, however, are much smaller, and the effects were not significant.

Reading times for the HiE and LoE versions of the texts were also analyzed. For each subject reading times for the sentences of the texts were regressed on number of syllables separately for the HiE and for the LoE texts. Two sets of regressions were carried out. The first included all sentences preceding the target sentences except sentence 1 (which tended to have much

longer reading times). The second included only those sentences unique to the HiE and LoE versions. It should be noted that the sentences unique to the HiE versions were intended to elicit expectations. No differences emerged in either analysis.

Finally, the relationship between the ratings for each text and the reading time for the target sentence was explored. For each of the four conditions, mean reading time for the primary target sentence for a text was regressed on the mean rating for that text version. A modest relationship emerged within the regression for the HiE-Unex condition. The slope in this regression was -348.4 msec., indicating that reading time decreased with increasing text rating ($t(22) = -1.94$, $p < .033$).

Discussion

The most important finding in this experiment is the effect of text version within the Unexpected condition. This finding suggests that an expectation becomes involved in the processing of the next sentence even when the expectation is unrelated to this sentence. Furthermore, the involvement of expectations can be rather costly; reading time for the primary target sentence in the HiE-Unex condition was increased by 316 msec. It makes reasonable the finding in Experiment 1 that a skilled reader does not generate expectations for every sentence. Given the costs involved, it is most efficient to generate an expectation only in the presence of sufficient constraints in the text on future events.

The overall pattern of results from Experiments 1 and 2 suggests that expectations can both help and hinder the integration of later sentences. Specifically, expectations help when they are correct and hinder when they are wrong. An expectation seems to acquire a privileged status which gives it priority in the processing of upcoming sentences, whether it is relevant to integrating these sentences or not. How this priority might be established is considered in the General Discussion.

The absence of significant effects in the secondary target sentences makes reasonable one form of the immediacy assumption of Just and Carpenter (1980). This is the assumption that processing for a particular sentence does not "overflow" to the following sentences. The pattern here suggests that integration processes for a sentence are carried as far as possible before going on to the next sentence. While the interaction pattern of the secondary target sentences was similar to that of the targets, no differences were significant. Thus, if an overflow exists, it seems to have a minor effect on processing time, especially in comparison to the effect of the manipulations on the processing time for the target sentences.

The fact that reading time in the HiE-Unex condition increased with strength of expectation (as measured by the rating task) lends further support to the claim that expectations are influencing reading time in this condition. This correlation seems to reflect the degree to which readers have committed

themselves to an expectation. The assumption is that the stronger the rating in the rating task, the more likely the subject in the reading task is to form an expectation about upcoming sentences, and the larger the interference when the expectation is not fulfilled.

One final result worth noting is that, contrary to other findings in the literature (Cirilo & Foss, 1980; Just & Carpenter, 1980), the important sentences did not consistently take longer to read. The Exp target sentences were always important next sentences; the Unex target sentences were always unimportant. The longest reading times were for unimportant sentences which happened to be unexpected (the HiE-Unex condition). Within the LoE text versions the important target sentences (Exp) were only marginally slower than the unimportant (Unex). This result is consistent with an analysis which suggests that the effect of importance on sentence reading time is mediated by expectations: sentences which are important and expected do not have long reading times.

EXPERIMENT 3

Throughout, the term "expectation" has been used to refer to two kinds of cognitive phenomena. A reader who has formed a general expectation has tagged certain text propositions as likely to be immediately followed up by the writer and possibly has activated some general elaborations on these propositions

(e.g., general causes and consequences). A reader who has formed a specific expectation has gone one step further and generated a specific prediction about upcoming content. If readers have a general expectation, then presumably they are set to encounter one of a range of possible next events in the narrative. In contrast if readers have a specific prediction, then they are set to encounter a single next event.

One reasonable model of expectation generation might be a two-step model in which the reader first generates a general expectation, and then further specifies the general expectation if possible. For example, for the train text the reader forms a general expectation that more will be said about the causes and/or consequences of the train screeching to a halt. The reader then goes on to generate a specific likely consequence based on prior text information (the fact that John is eating soup).

There are at least two reasons why readers might not be regularly going on to form specific predictions once a general expectation is generated. First, it seems to be a general property of good narratives that the specifics cannot be predicted. Stories in which the reader could predict exactly what will happen next would be extremely dull. It is unlikely that good readers would adopt a strategy of regularly generating specific predictions when these predictions are likely to be wrong. Second, the generation of a specific prediction is likely to require a fair amount of processing time. In the train text

for example, it requires a search of the text for a likely object to be incorporated into the expectation and an evaluation of the candidates found (e.g., waiter, John, soup, etc.). The literature on antecedent search suggests that the search for and evaluation of a prior argument in a text can be time-consuming (Cirilo, 1981; Clark & Sengul, 1979; Garrod & Sanford, 1977). If this is the case, then it seems to be a poor strategy for the reader to adopt. Why slow down to predict what will happen next rather than simply reading on to find out what actually did happen next?

For the texts used in Experiments 1 and 2 it was possible for the reader to generate a specific prediction for the NIE versionn. Experiment 3 tested the hypothesis that readers in the first two experiments were actually forming these highly specific predictions. The specific prediction elicited by the NIE texts in Experiment 2 always involved a target argument introduced earlier in the portion of the text that was common to the NIE and LoE versions (e.g., "soup" in the train text). Experiment 3 probed the availability of that target argument. If subjects actually generate the specific prediction while reading, then the target argument should be retrieved and should be held in a highly available form as part of the prediction. As a result, target arguments which form part of a specific prediction should be more available than they would otherwise be. In the absence of predictive processes, the availability of an argument has been shown to vary with the distance of the last mention in the text

(Carpenter & Just, 1977, 1978; Cirilo, 1981; Clark & Sengul, 1979; Lesgold et al, 1979). In contrast if a prediction involving the argument has been made, then distance of last mention should not have an effect on availability. The argument should be highly available regardless of distance.

In this experiment each text had four versions. The four versions of the train text are presented in Table 5. Two factors which should affect the availability of a target argument (e.g., soup) were varied orthogonally: the distance of the argument's last mention in the text, and the degree to which the reader has a specific prediction involving the target argument.

Insert Table 5 about here

Availability was measured using a forced-choice task. Subjects read the texts one sentence at a time. They were interrupted at the point where an expectation should be generated for the NIE version of the text. The target word was presented along with a distracter. Subjects indicated which word had appeared in the text. If the availability hypotheses are correct, then response time for the target arguments for the NIE texts should be faster than to the LoE. Furthermore, distance should affect response time for the target argument for the LoE texts, with the distant target requiring more time than the close. No distance effect (or a reduced distance effect) should be found for the NIE texts. Thus distance and text version

should interact in their effect on response time for the target argument.

Methods

Subjects. Twenty-six subjects from the University of Michigan community participated. Two subjects were excluded for failure to follow instructions. Subjects were paid $3.50 for 30 to 40 minutes of participation.

Materials. The texts used in Experiment 2 were used as the HiE-Close and LoE-Close versions of Experiment 3. In all of these texts the target argument appeared in the third to last sentence of the text. To create the Distant versions of each text, two filler sentences were inserted somewhere between the last mention of the argument and the end of the text. These filler sentences were identical for the HiE and LoE versions of a given text. They did not refer to the target argument. Three sentences back was chosen as the distance in the Close condition because pilot studies suggested that at a distance of two sentences back the argument might still reside in verbatim memory. The Close and Distant versions of the train text are given in Table 5.

The HiE version of each text was designed to elicit a specific content prediction involving the target argument. Care was taken to insure that the expectation sentences in the HiE version of a text were not highly semantically related to the target argument. For example, "soup" is not semantically related

to a train screeching to a stop except in the context of the full train text. As a result, the target argument could not have been made available in the HiE version by activated semantic pathways rather than by predictive processes operating on the text.

All four versions of the texts were validated in the same rating task described in Experiment 2. Counts were made of the number of subjects who included the target word as part of their prediction of what should happen next. Texts were rewritten to insure that the HiE versions of each text always elicited mention of the target word more frequently than the LoE versions. The final mean percentages of subjects mentioning the target argument were 90 for the HiE-Close, 87 for the HiE-Distant, 55 for the LoE-Close and 53 for the LoE-Distant. An ANOVA on the mean percentages for each text revealed a significant effect of text version ($F(1,23) = 97.08$, $p < .001$); neither distance nor the interaction of distance with text version approached significance.

For each text, a distracter word was chosen to be paired with the target argument for the forced-choice task. The distracters were words which had not appeared in the text. They were always the same length (in letters) and of similar word frequency as the target word. For the filler texts, similar pairs were constructed with one member coming from the text and one not.

The forced-choice procedure was adopted after running a pilot experiment in which subjects were asked to make an old-new

judgment for a single target word. This old-new probe task was similar to a procedure used by McKoon and Ratcliff (1980; Dell, McKoon & Ratcliff, 1983). The single-word probe task was abandoned because it overemphasized memory for the exact word. Some subjects reported making errors when they were unsure whether a target word or a synonym had actually appeared in the text. The distracters in the forced-choice task were chosen to be clearly unrelated to the text to reduce this problem.

Design. Each subject read one version of each of the 24 texts. There were four experimental conditions formed by the crossing of text version (MiE vs. LoE) with distance (Close vs. Distant). A given subject saw six texts in each of the four conditions. Across the whole experiment, each text and target word appeared equally often in each condition. Order of presentation of texts was randomized every four subjects.

Procedure. Subjects read a total of 84 texts. The first 36 were practice trials. The remaining 48 were the 24 experimental plus 24 fillers.

Subjects read each text one sentence at a time on a CRT. A press of the space bar erased the current sentence and brought on the next. At some point in each text the next sentence did not appear when the space bar was pressed. Instead, a large down-pointing arrow (5 cms. high) appeared on the left side of the screen. After a 333 msec. delay the test words appeared under the arrow. The words were displayed on the same line, separated by three character spaces. Subjects responded by pressing the

left ("z") or right ("/") response key to indicate which word had appeared in the text.

The target word from the experimental texts always appeared as the lefthand word of the pair. This was to increase the likelihood that subjects were actually reading the target word from the experimental texts rather than responding on the basis of the distracter. Across the whole experiment, the correct response appeared equally often on the left and on the right.

After the subject responded to the test words, a feedback message was displayed. After one second, the message was automatically replaced by a verification sentence. Subjects indicated whether or not the sentence was true of the text by pressing one of the two response keys. Again, feedback was displayed for one second.

Results

Means were computed as in Experiments 1 and 2. Error trials were excluded from the analysis. Mean response times and error rates are displayed in Table 6. The means were submitted to two ANOVAs. In the ANOVA by subjects, subjects were nested within group; both factors were crossed with text version and distance. In the ANOVA by text, target words were nested within text group; both factors were crossed with text version and distance. The only significant effect was that of distance ($F_1(1,23) = 5.61$, $p < .05$, $MSe = 12712$; $F_2(1,20) = 4.92$, $p < .05$, $MSe = 12185$).

Insert Table 6 about here

Although the means do exhibit the predicted interaction pattern, a closer inspection of the data supports the statistical analysis. The large mean response time in the LoE-Distant cell is due to one rather variable subject who made two slow responses in this condition. If this subject is excluded from the analysis, the mean for the LoE-Distant cell drops to 1009 msec.[1]

Discussion

The lack of an effect of text version or of an interaction between text version and distance suggests that the target argument was no more available at the probe point for the NIE texts than for the LoE. These findings provide no support for the claim that subjects were forming specific predictions as they read the NIE texts.

The finding of a significant distance effect indicates that the forced-choice task was sensitive enough to detect differences in the time needed to access the text representation. This finding is consistent with the results of reading time studies which suggest that antecedent search time varies with the distance of the antecedent.

There are several possible accounts of the data. Perhaps the least interesting would be an account which claims that subjects were forming specific predictions all of the time and

were equally likely to have formed the correct prediction (i.e., involving the target word) for both the NIE and LoE texts. The results of the rating task render this account unlikely. In that task subjects were much more likely to use the target noun in a prediction for the NIE text versions than for the LoE versions.

A second possibility is that subjects did not have enough time to form a specific prediction. The processing hypothesized for generating a specific prediction included the process of retrieving the specific argument from the text. As the data suggest, this is a time-consuming process in the distant condition. It may be that predictive activities are initiated at the end of the sentence as the subject presses the key to go on to the next sentence. As a result, when the forced-choice probe is presented, the specific prediction has not yet been formed. Analyses in Experiments 1 and 2 found no differences in time spent reading the NIE vs. LoE texts themselves. This result supports the claim that readers were not taking the extra time to generate a specific prediction before going on to the next sentence.

An objection can be raised to the claim that readers generate a specific prediction as they move on to the next sentence. Time to make a prediction can be expected to vary from text to text, depending in part on how obvious the prediction is. Thus, for each subject the mean response time for the NIE probes should be a mix of trials on which the argument had been retrieved to form the specific prediction and trials on which it

had not been retrieved before the probe was presented. This interpretation still predicts that the HiE means should have been faster than the LoE. The data do not support this prediction.

It is clear from the rating task that readers can form the specific predictions elicited by the HiE texts when given unlimited time and when instructed to generate a prediction. Yet it is also clear from the forced-choice data that in the course of the reading task subjects were not generating such specific predictions before going on to read the next sentence. It is certainly possible that if the forced-choice probe had been delayed by several seconds, evidence of the presence of specific predictions might have been discovered. But if specific predictions are so time-consuming to generate, then it is unlikely that such predictions could have played a role in the first two experiments. In those experiments the sentence which fulfilled the prediction appeared immediately after the sentence which elicited the prediction. The effects in those experiments would therefore seem likely to reflect the influence of general expectations rather than highly specified predictions.

The lack of any evidence that readers regularly form highly specific predictions suggests that the usefulness of expectations is not limited to highly predictable texts. Most narratives are not so transparent that the reader can predict exactly what will happen next. The results suggest that general expectations may be useful in reading a broad range of narratives, unpredictable as well as predictable.

General Discussion

The three experiments reported here were designed to examine the role of expectations in sentence integration. In the first two experiments, subjects were faster both to read and to judge an expectation-fulfilling target sentence when it followed the appropriate HiE text version than when it followed the LoE. Subjects were also faster to judge a target sentence to be unrelated when it followed a HiE text. They were slower to read a related target sentence which violated the expectation when the sentence followed a HiE text. In the third experiment, subjects were no faster to respond to a target word which followed a HiE text than when it followed a LoE, even though the target word was part of the specific prediction elicited by the HiE text in a separate rating task.

The results of these experiments provide support for an expectation model of the following sort. Readers regularly generate expectations about upcoming events as they read narratives. These expectations consist of a special subset of the text propositions which the reader has tagged as likely to be followed up by the writer, and possibly some general causes and/or consequences for the tagged information, activated in long term memory. These expectations are not generated for every sentence in every text; they are generated for those sentences which are perceived to be causally important to the narrative.

Thus, at some points in a text readers may have strong expectations about upcoming events, and at other points they may have none. Having an expectation about an immediately upcoming event has consequences for the integration of upcoming sentences. When an expectation is present, readers try to relate the next sentence to the contents of the expectation. When no expectation is present, readers may have to search the prior text for related information. A correct expectation helps the integration process by allowing the reader to bypass some search. An incorrect expectation interferes by postponing a necessary search process.

Once an expectation is formed, it seems to be given priority in the search for causal links. This raises the question of what mechanism might be proposed to confer this priority status on the subset of text information which is tagged in the expectation. In current models a limited-capacity working memory is used to give priority to a subset of the text information (Kintsch & van Dijk, 1978; Lesgold et al, 1979). The assumption is that the contents of working memory are searched first for links to the current sentence. The search goes beyond working memory only when the search for links to information in working memory fails. It might seem reasonable, then, to claim than an expectation is always maintained in working memory, and thus has priority in the search.

The working memory account, however, has difficulty with the Unrelated results in Experiment 1. Even in the absence of an expectation, working memory is assumed to contain a subset of

text information. If subjects are fast in the HiE-Unrelated condition because they are simply checking the Unrelated target against the contents of working memory (where the expectation resides), then they should be equally fast in responding to the LoE-Unrelated condition.

An alternative might be to hypothesize that the reader is arranging the pieces of the causal chain of the story in a separate workspace (Olson, Duffy & Mack, in press). Elements in the causal chain are thus given special status apart from the status given to recent propositions currently residing in working memory. Such a claim is embodied in the concept of the macro-structure which is constructed during reading in the Kintsch and van Dijk (1978) model and in the "weird list" in the model of Schank and Abelson (1977). In both cases a subset of all the information presented is identified as useful for understanding the causal backbone of the narrative. When the reader forms an expectation about immediately upcoming events, one outcome may be the allocation of processing priority to the parts of the causal chain involved in the expectation. This results in a commitment to spend time trying to link the next sentence to the prioritized propositions. Such a commitment can account for the Unrelated results noted above. In the HiE-Unrelated condition, readers have identified some text information as part of the causal chain and as likely to be followed up. This information is placed in the workspace and is used to make a judgment about the Unrelated target sentence. In the LoE-Unrelated condition, the reader has

not placed causal chain information in the workspace. The judgment of the target sentence thus requires a search of the whole text rather than a search of the subset of information in the workspace.

Conclusion

One major goal of research on reading comprehension has been to characterize the search and inference processes involved in integrating the current sentence with the text propositions which have already been read. Processing is optimized to the extent that the reader can focus the integration attempt on the relevant subset of the prior text propositions. In current models working memory is used to give priority in antecedent search to a recent subset of the already-read text propositions. The experiments reported here suggest that expectations are another way that readers confer priority on relevant prior text information, information which will be especially useful for causal inferencing.

Reference Notes

1. Keenan, J. Inferring causal connections in prose comprehension. Paper presented at the Annual Meeting of the American Psychological Association, Toronto, August 1978.
2. Wilensky, R. Points: A theory of story content. Technical Report M80/17. Berkeley, Ca.: Electronics Research Lab, University of California at Berkeley, 1980.
3. Rumelhart, D. Understanding understanding. Center for Human Information Processing Report No. 100. University of California, San Diego, January 1981.

N.J.: Erlbaum, 1980.

Dell, G., McKoon, G., & Ratcliff, R. The activation of Antecedent information during the processing of anaphoric reference in reading. Journal of Verbal Learning and Verbal Behavior, 1983, 22, 121-132.

Ehrlich, S. & Rayner, K. Contextual effects on word perception and eye movements during reading. Journal of Verbal Learning and Verbal Behavior, 1981, 20, 641-655.

Eisenberg, P. & Becker, C. Semantic context effects in visual word recognition, sentence processing, and reading: Evidence for semantic strategies. Journal of Experimental Psychology: Human Perception and Performance, 1982, 8, 739-756.

Fahlman, S. NETL: A system for representing and using real-world knowledge. Cambridge, Ma.: MIT Press, 1979.

Fodor, J., Bever, T., & Garrett, M. The Psychology of language. New York: McGraw-Hill, 1974.

Garrod, S. & Sanford, A. Interpreting anaphoric relations: The integration of semantic information while reading. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 77-90.

Goodman, K. Reading: A psycholinguistic guessing game. Journal of the Reading Specialist, 1967, 6, 126-135.

Haber, R. Visual perception. Annual Review of Psychology, 1978, 29, 31-59.

Just, M. & Carpenter, P. A theory of reading: From eye fixations to comprehension. Psychological Review, 1980, 87, 329-354.

References

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

Carpenter, P. & Just, M. Reading comprehension as eyes see it. In M. Just & P. Carpenter (Eds.), Cognitive Processes in Comprehension. Hilledale, N. J.: Erlbaum, 1977.

Carpenter, P. & Just, M. Integrative processes in comprehension. In D. LaBerge & S. Samuels (Eds.), Basic Processes in reading: Perception and comprehension. Hilledale, N. J.: Erlbaum, 1978.

Cirilo, R. Referential coherence and text structure in story comprehension. Journal of Verbal Learning and Verbal Behavior, 1981, 20, 358-376.

Cirilo, R. & Foss, D. Text structure and reading time for sentences. Journal of Verbal Learning and Verbal Behavior, 1980, 19, 96-109.

Clark, H. & Haviland, S. Comprehension and the given-new contract. In R. Freedle (Ed.), Discourse comprehension and production. Norwood, N. J.: Ablex, 1977.

Clark, H. & Sengul, C. In search of referents for nouns and pronouns. Memory and Cognition, 1979, 7, 35-41.

Collins, A., Brown, J., & Larkin, K. Inference in text understanding. In R. Spiro, B. Bruce, & W. Brewer, (Eds.), Theoretical issues in reading comprehension. Hilledale,

Kintsch, W. & van Dijk, T. Toward a model of text comprehension and production. Psychological Review, 1978, 85, 363-394.

Longold, A., Roth, S. & Curtis, M. Foregrounding effects in discourse comprehension. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 291-308.

McClelland, J. & O'Regan, J. Expectations increase the benefit derived from parafoveal visual information in reading words aloud. Journal of Experimental Psychology: Human Perception and Performance, 1981, 7, 634-644.

McKoon, G. & Ratcliff, R. The comprehension processes and memory structures involved in anaphoric reference. Journal of Verbal Learning and Verbal Behavior, 1980, 19, 668-682.

Mitchell, D. & Green, D. The effects of context and content on immediate processing in reading. Quarterly Journal of Experimental Psychology, 1978, 30, 609-636.

Olson, G., Duffy, S. & Mack, R. Thinking-out-loud as a method for studying real-time comprehension processes. In D. Kieras & M. Just (Eds.), New methods in the study of immediate processes in comprehension. Hillsdale, N. J.: Erlbaum, in press.

Schank, R. & Abelson, R. Scripts, plans, goals and understanding. Hillsdale, N. J.: Erlbaum, 1977.

Smith, F. Understanding reading. New York: Holt, Rinehart & Winston, 1971.

Footnotes

This research was supported by a research grant from the National Institute of Education (NIE G-79-0133) to Gary M. Olson; the paper was completed while the author held a NIH Postdoctoral Traineeship at the University of Massachusetts. Portions of this research were presented at the Annual Meeting of the Eastern Psychological Association, Philadelphia, April 1983. This report is based on a doctoral dissertation submitted to the University of Michigan. I am grateful to the members of my committee, Gary Olson (Chair), Keith Holyoak, J. E. Keith Smith, and Eric Rabkin, for their guidance throughout the research. I would also like to thank Jerry Myers and Keith Rayner for thoughtful comments on an earlier draft of the paper, and Don Fisher for helpful suggestions throughout the course of the research. Requests for reprints should be sent to Susan Duffy, Psychology Department, Tobin Hall, University of Massachusetts, Amherst, Ma. 01003.

[1]The same pattern of results was obtained in the pilot experiment using the single word probe task instead of a forced-choice task. Subjects were always probed with the target argument for the experimental texts; their task was to indicate whether the word had appeared in the text. Mean response times for 13 subjects were (with error percentages in parentheses): 994 (6.4) in the HIE-Close condition; 1057 (12.8) in the HIE-Distant; 979 (6.4) in the LoE-Close; and 1034 (11.5) in the LoE-Distant.

Table 1

Example Text Versions from Experiment 1

High Expectation Version:

John was eating his first meal ever in the dining car of a train.

The waiter brought him a large bowl of bean soup.

John tasted the hot soup carefully.

Suddenly the train screeched to a stop.

Low Expectation Version:

John was eating his first meal ever in the dining car of a train.

The waiter brought him a large bowl of bean soup.

John tasted the hot soup carefully.

The train began to slow down entering a station.

TARGET SENTENCES FOR BOTH VERSIONS:

Related: The hot soup spilled into John's lap.

Unrelated: That night the whole forest burned down.

Table 2

Mean Response Times (in msec) and Percent Disagreement for the Target Sentences in Experiment 1

	Text Version	
Target Sentence	HiE	LoE
Related	1588 (2)	2044 (13)
Unrelated	1799 (7)	1926 (5)

Note: Disagreement percents are in parentheses.

Table 3

Example Text Versions from Experiment 2

High Expectation Version:

John was eating his first meal ever in the dining car of a train.

The waiter brought him a large bowl of bean soup.

John tasted the hot soup carefully.

He reached for the salt shaker.

Suddenly the train screeched to a stop.

Low Expectation Version:

John was eating his first meal ever in the dining car of a train.

The waiter brought him a large bowl of bean soup.

John tasted the hot soup carefully.

He reached for the salt shaker.

The train slowed down entering a station.

TARGET SENTENCES FOR BOTH VERSIONS:

Expected: The soup spilled all over John's clean shirt and pants.

His paper napkin was no help.

Unexpected: The waiter came to offer John some coffee.

John said he would like tea with cream.

Table 4

Mean Response Times (in msec) for Target Sentences in Experiment 2

		Text Version	
		HiE	LoE
TARGET SENTENCE			
Target Sentence Type	Exp	1543	1711
	Unex	1879	1563
SECONDARY TARGET SENTENCE			
Target Sentence Type	Exp	1406	1425
	Unex	1467	1316

Table 5

Example Text Versions from Experiment 3

High Expectation - Close

John was eating his first meal ever in the dining car of a train.
The waiter brought him a large bowl of bean soup.
John tasted the hot soup carefully.
He reached for the salt shaker.
Suddenly the train screeched to a stop.

Low Expectation - Close

John was eating his first meal ever in the dining car of a train.
The waiter brought him a large bowl of bean soup.
John tasted the hot soup carefully.
He reached for the salt shaker.
The train slowed down entering a station.

High Expectation - Distant

John was eating his first meal ever in the dining car of a train.
The waiter brought him a large bowl of bean soup.
John tasted the hot soup carefully.
He reached for the salt shaker.
It seemed to be made of sterling silver.

The railroad's initials were engraved on one side.
Suddenly the train screeched to a stop.

Low Expectation - Distant

John was eating his first meal ever in the dining car of a train.
The waiter brought him a large bowl of bean soup.
John tasted the hot soup carefully.
He reached for the salt shaker.
It seemed to be made of sterling silver.
The railroad's initials were engraved on one side
The train slowed down entering a station.

TARGET WORD FOR ALL VERSIONS: soup

Table 6

Mean Response Times (in msec) and Percent Error for Experiment 3

		Text Version	
		HiE	LoE
Distance	Close	960 (1.4)	972 (0.7)
	Distant	1009 (0.7)	1033 (0.7)

Note: Error percents are in parentheses.