# CANADIAN THESES

# THÈSES CANADIENNES

## NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

## AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

# COMPREHENSION AND QUESTION ANSWERING:

# A COMPARATIVE STUDY



A Thesis Submitted
to the Faculty of Graduate Studies and Research
in Partial Fulfillment of the Requirements for the Degree of
Master of Arts

McGill University
Department of Educational Psychology

Montréal, Québec

# ABSTRACT

This study was designed to investigate the extent to which the cognitive structures and processes underlying language comprehension are related to those involved in question answering. In addition, it examined the relationship between comprehension and question answering in the presence of text. The latter task is frequently used for the assessment of comprehension. As well, the study included two experimental variables suggested by current theories of discourse processing: (a) Type of question (text-based and frame-based questions) and, (b) type of text structure (narrative and descriptive). Effects of these experimental variables on propositions recalled or inferred in response to questions and in free recall were studied. The latter was assumed to reflect comprehension.

Analyses of data revealed that comprehension contributes significantly to the general level of performance in answering questions. The pattern of results suggested that question answering in the absence of text depends on the availability of relevant information in memory representation for text. Answering questions while referring to text, on the other hand, was found to involve comprehension processes as well as question-specific processes. Moreover, this task was found to produce highly literal responses. These findings were interpreted as evidence for the superiority of the former task as a measure of

comprehension.

Results also indicated the relative difficulty of descriptive discourse over narratives and that of questions requiring the organization of information at a global level over those interrogating local-level information. The two types of questions were found to elicit different processes that are text-structure dependent. The educational and theoretical implications of these findings are discussed.

## RESUME

Notre étude a eu pour but d'examiner dans quelle mesure les structures et processus cognitifs sous-jacents à la compréhension de texte sont reliés aux structures et processus cognitifs impliqués lors des réponses apportées aux questions basées sur le texte. De plus, nous avons étudié la relation entre la compréhension et les réponses aux questions, ce, à partir du texte. Ce genre de tâche à partir d'un texte sert fréquemment à l'évaluation de la compréhension.

Dans notre recherche, nous avons utilisé deux types de questions. En conceptualisant les problèmes empiriques à l'aide d'un modèle théorique de traitement du discours et de l'information, nous avons observé l'effect produit par la structure du texte sur les propositions remémorées ou inférées lors des réponses apportées aux questions posées de même que lors du rappel libre; le rappel libre reflétant supposément le degré de compréhension.

L'analyse des données a révélé que la compréhension contribue significativement au niveau de performance de réponse aux questions. La structure des résultats obtenus a démontré que la capacité de répondre à des questions en l'absence du texte dépend de l'information pertinente dont le sujet dispose dans sa représentation cognitive d'un texte. Nous avons également trouvé que répondre à des questions en présence du texte nécessite, en

plus des processus généraux de compréhension, des processus spécifiques de réponse aux questions. Cette dernière tâche favorise la production de réponses littérales. Ceci nous amène donc à penser que la tâche de répondre aux questions en l'absence du texte est une meilleure tâche d'évaluation de la compréhension que celle de répondre aux questions à l'aide du texte.

Les résultats nous ont aussi indiqué la relative difficulté du discours descriptif par rapport au discours narratif de même que des questions exigeant une organisation globale par rapport aux questions conduisant à une information plus circonscrite dans le texte. Nous avons constaté que ces deux types de questions font appel à différents processus cognitifs liés à la structure du texte.

Nous avons finalement discuté des implications pédagogiques et théoriques que ces données soulévent.

# ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Carl H. Frederiksen, who introduced me to research and thinking about thinking. Thanks to him this process became an invaluable learning experience and I am grateful to him for providing me the opportunity of working in his discourse processing laboratory. I also wish to thank all the people of the lab who made this process an enjoyable one and I am grateful to Dr. Janet Donin Frederiksen for teaching me techniques of discourse analysis.

I am particularily indebted to Dr. Eigil Pedersen for his extremely constructive help on personal, practical, and academic levels, at various stages of this research.

To Lilli Kormos go special thanks for her support. Selflessly, she shared with me her research experience and I have benefitted greatly from discussions with her at every single stage of this study.

I would also like to acknowledge the help of Donal MacAnaney for spending time with me discussing methodological problems and that of Christine Laphkas who, together with Dr. Pedersen, provided me access to private schools during the strike of Québec's teachers in the winter of 1983, and helped me in collecting data.

Finally, very special thanks go to my husband, Massoud Pourafzal, for encouraging me to learn more and more by doing so

himself, for often "putting the default values into question", and ... for everything else.

# TABLE OF CONTENTS

# TABLES

# FIGURES

# CHAPTER I

# INTRODUCTION

What do we do when we wonder whether someone has understood something? We ask him or her questions.

Educators and the rest of us assess comprehension by asking questions. Psychological researchers, on the other hand, have tended to favour recall protocols as measures of comprehension, because, as compared to question-answering tasks, presumably fewer constraints operate in free recall.

Does the difference in the degree of constraint imposed on cognitive processes make the two types of task largely independent? As a matter of fact, what do people do cognitively when they understand a passage and reconstruct it in a recall task? What do they do when they answer questions about it? To what extent can we predict the occurence of answers by knowledge of the cognitive structures and processes that people exhibit spontaneously in free recall? Is the ability to answer questions distinct from the ability to comprehend? Is there a special competency related to answering questions or is it the case that once we have understood an organized unit of information we can answer the comprehension questions relevant to it? Will a comparative study of comprehension and question answering reveal

differences interpretable in terms of competence-performance or comprehension-production discrepancies? In other words, do the processes that people do exhibit in an ambiguously structured task like free recall predict those that they can exhibit when provided with explicit demands for specific processes?

Many teachers have heard students complain that they could not answer test questions despite having understood the lesson. Sometimes students also feel that they can answer questions quite accurately without having 'really' understood the material. To what extent can we empirically support these subjective experiences of a certain independence between comprehension and question answering?

Psychologists who study question answering from a cognitive perspective also have conceptualized comprehension and question answering as non-identical processes. How could they be? The mere presentation of questions to the comprehender is a new elicitation condition introducing processing demands other than those imposed by the to-be-comprehended information itself. However, how closely the processes involved in the two types of task are related remains an unanswered question.

The central purpose of the research reported here was to investigate some of the issues raised above. Grade-six students of eleven to twelve years of age were asked to read texts, recall them and answer comprehension questions relevant to them. Since comprehension is often assessed by having students answer questions while the reading material is available to them,

subjects were also asked to answer questions in the presence of text. The main objective was to examine the degree to which the three tasks, namely, free recall, question answering, and question answering with access to text, are related. Throughout the research, it was assumed that recall protocols reflect students' comprehension and that answers reflect their question answering ability (see Chapter 3 for the rationale for this assumption).

The study also examines the effect of type of text structure (e.g., narrative versus descriptive) on question answering and comprehension. In addition, it investigates some effects that varying the cognitive demands of questions has on question answering.

In view of the fact that questioning is the primary method of educational evaluation in general and assessment of comprehension in particular, knowledge about question answering and its relationship to comprehension has obvious implications for education. Moreover, fundamental research on the cognitive mechanisms that underlie question answering can provide basic information about human cognitive structures and processes. This study is an attempt to improve our understanding of the question-answering process especially as it relates to language understanding.

# CHAPTER II

# QUESTION ANSWERING

# AND

# DISCOURSE COMPREHENSION

There exists a vast and varied literature on question answering. In an article entitled The multidisciplinary study of questioning, Dillon (1982) refers to several fields which are concerned with the processes of question answering. These include philosophy, linguistics, logic, cognitive psychology, anthropology, sociolinguistics, education, psychotherapy, and others. The present review is inevitably highly selective and the selection is based on a criterion of presumed relevance to the study.

This review is presented in two sections. The first section concerns the more application-oriented investigations, mainly conducted by educational researchers. The second section reviews the more 'basic' and theoretical literature, stemming mainly from the work of cognitive psychologists. It deals with theory and research related to discourse processing, especially language comprehension, and with information-processing studies of question answering.

Question answering is not only an important means of conveying information but it is also an important process through which learning may be assessed and enhanced. Several areas in education are concerned with the study of questioning. The entire area of educational evaluation, measurement, testing and assessment bears upon question answering. Moreover, traditionally, asking questions has been considered a teaching technique and, thus, teacher questions as well as student self-interrogation have been examined by educators. These studies have only a marginal relevance to the present research and will not be reviewed in detail, although some reference to assessment and testing notions will be made in the context of different question types.

However, there is a voluminous literature on the effects of questions on 'prose learning'. Because it deals with connected discourse materials, this literature, which is often referred to as the 'adjunct questions' literature, is more relevant to the purposes of the present study and is reviewed below.

## Effects of Adjunct Questions

### The Experimental Paradigm

Concerned with the improvement of student learning and retention from text, educational researchers have been studying the effects of various types of study aids such as advance organizers (Ausubel, 1960, 1965; Ausubel & Fitzgerald, 1965; Barnes & Clawson, 1975) or adjunct questions (Frase, 1967, 1968a, 1968b; Rothkopf, 1966; Rothkopf & Bisbicos, 1967), to determine

the extent of their instructional usefulness (for reviews, see T. H. Anderson, 1980; Rickards, 1980).

Although a number of early investigations (e.g., Jones, 1923) had examined the facilitative effects of questioning, and despite the fact that, for decades, teaching-methods books have recommended questioning techniques (e.g., the SQ3R method designed by F. P. Robinson, 1970; or the PQ4R technique of Thomas & H. A. Robinson, 1977), it is the work of Rothkopf (1965, 1966; Rothkopf & Bisbicos, 1967) that has stimulated the numerous, more recent, studies on the learning consequences of asking students questions about what they are reading (see R. C. Anderson & Biddle, 1975 for a detailed review; see also Andre, 1979; McConkie, 1977; Reder, 1980; Rickards, 1979).

An experimental paradigm similar to that designed by Rothkopf (1966) characterizes the adjunct questions research area: Students are asked to answer questions placed either before or after textual material they are reading. Later, they are given a criterion test involving the same questions as before, new questions, or both. The objective is to determine the effects of the initial questioning on posttest performance.

While numerous inconsistencies have been reported in the literature, it is safe to state that, compared to a reading-only control group, facilitative effects have been found not only when the posttest consists of repeated questions ('direct' or 'intentional' effects), but also when it involves new items ('indirect' or 'incidental' effects) (Anderson & Biddle, 1975).

Rothkopf's (1966) experiment demonstrated both types of

effects; the indirect effects were, however, limited to the case where the adjunct questions appeared after the relevant reading material. In fact, many of the inconsistencies in the findings of questioning studies concern the indirect effects (Anderson & Biddle, 1975). It seems however that, as found by Rothkopf (1966; Rothkopf & Bisbicos, 1967), in most studies, prequestioning does not facilitate and can even inhibit performance on new posttest items (Anderson & Biddle, 1975). Retention of incidental information is facilitated by questions that follow reading passages (e.g., Frase, 1967, 1968a).

Besides the position of the initial items and the degree of novelty of the posttest, adjunct questions studies have typically included such independent variables as frequency of questioning (e.g., Frase, 1968b), whether or not feedback on the adjunct questions is given (e.g., Frase, 1967; Friedman & Rickards, 1981), or the type of the questions (see Andre, 1979 for a review).

The Differential Effects of Various Types of Questions

With the recent development of discourse analysis techniques, free recall tasks are increasingly used for the assessment of learning (Johnston, 1981; McConkie, 1977). In the past, techniques such as the 'cloze' procedure (Taylor, 1953) - i.e., asking the student to fill the often randomly selected blanks of a passage- have been used as a measure of reading comprehension (Collins, Brown, Morgan & Brewer, 1977). However, question-oriented techniques are the ones that are used most often (McConkie, 1977).

Yet, the classification of different types of questions has been and (despite attempts to develop explicit rule-governed procedures for objective question generation and classification (e.g., R. C. Anderson, 1972; Bormuth, 1970; Finn, 1975)) continues to be an unresolved problem for questioning research and educational assessment (Carrier & Fautsch-Patridge, 1981; Johnston, 1981, 1982; McConkie, 1977).

Although most adjunct questions investigators have used 'factual' questions only, and many others have classified questions subjectively and/or atheoretically (see Carrier et al. 1981 for a discussion of this problem), a number of frameworks have been used for generating various types of questions.

One such framework is Bloom's taxonomy of educational objectives which includes the following cognitive categories: knowledge, comprehension, application, analysis, synthesis, and evaluation (Bloom, Engelhart, Furst, Hill & Krathwohl, 1956). This taxonomy was used by Hunkins (1969), among others. Hunkins' study compared the effects of 'analysis' and 'evaluation' questions with those of 'knowledge' questions. The posttest consisted of six sub-scores, one for each of Bloom's categories. Reading ability was a significant predictor of all six scores. Moreover, children who received the analysis and evaluation questions did significantly better than the knowledge-questions students on the evaluation subtest but not on the five 'lower' subtests.

Somewhat more surprising results have been found by Shavelson, Berliner, Ravitch and Loeding (1974). In this study,

questions were either of 'lower order', corresponding to Bloom's 'knowledge' category, or of 'higher order' relevant to the taxonomy's 'comprehension', 'application' and 'analysis' categories. On a posttest, the group that received higher-order questions placed after the passage performed better than all other experimental groups but, surprisingly, not better than the no-question control group.

Several authors (e.g., R. C. Anderson, 1972; Baker, 1974; Bormuth, 1970) have questioned the adequacy of Bloom's taxonomy for classifying test items and Andre (1979) has pointed out the vagueness of the classification system used in both of the above studies and has attributed some of the inconsistencies to the low reliability for the assignment of questions to categories. However, Bloom's taxonomy together with the concern for behaviorally stated objectives (Mager, 1962) still influence educational testing and measurement (see e.g., Bloom, Hastings & Madaus, 1971; Green, 1975; Lindvall, 1975; TenBrink, 1974).

In his 1963 paper, Glaser made a distinction between norm-referenced and criterion-referenced achievement measures. The former aims at the relative ranking of students with respect to their test performance. Test items are, therefore, selected on the basis of their discriminating power (Millman, 1974). Glaser advocated the use of criterion-referenced tests, i.e., measures that provide "explicit information as to what the individual can or cannot do" (1963, p. 520).

In psychometric work such as that suggested by Glaser, concern for the validity and interpretability rather than mere

reliability of test items has increased. Hence, there is a greater emphasis upon the explicit specification of the content, format, and selection of the items (Millman, 1974).

To avoid misinterpretation of the term 'criterion-referenced testing', Hively (1974) replaced it with that of 'domain-referenced testing'. "A 'domain' may consist of any clearly specified set of items" (p. 10). It is a well-defined set of tasks from which test items are randomly selected (Millman, 1974).

The concern for systematicity in the construction and validation of test items (T. H. Anderson, Wardrop, Hively, Muller, R. I. Anderson, Hastings & J. Frederiksen, 1977) is reflected in the works of authors who propose algorithmic approaches to the generation of questions. As McConkie (1977) put it, "the test questions used are then a sample of all questions that could be generated by the same algorithm, thus making clear the domain from which they came" (p. 9).

Strongly criticizing the subjective nature of question generation procedures, Bormuth (1970) has advocated the use of operational definitions for deriving a population of criterion-referenced test items from a given instructional statement, so that, by applying the operations, two independent test writers generate the same items. The definition would include a set of operations through which the syntactic structure of the sentences is analyzed, and a set of transformations to produce test items. Using notions of transformational generative grammar, Bormuth has designed such item-producing rules and has used them in a study

(Bormuth, Manning, Carr & Pearson, 1970) which examined children's comprehension of various syntactic structures and found that accuracy of response was affected by the type of question asked.

Bormuth's proposal is, however, heavily syntax-based and as such it is not directly useful for the assessment of the learner's understanding of the meaning of a passage. In an attempt to overcome this problem, R. C. Anderson (1972) has extended Bormuth's work and has provided a number of additional transformations to generate test items.

The scheme proposed by Anderson is aimed at distinguishing comprehension from non-comprehension. Emphasizing the importance of semantic, as compared to orthographic and phonological, encoding in comprehension, the model focuses on the relationship of the question's wording to that of the instruction, and describes procedures for generating 'verbatim' questions, 'paraphrase' questions, and questions formed by substituting more general or more particular terms for the original terms.

A number of studies have examined the effects of adjunct questions that vary according to this framework. Somewhat related to Anderson's concern for comprehension are the studies that conceptualize question type in terms of levels of processing (Craik & Lockhart, 1972). From this standpoint it would be expected that adjunct questions that presumably require 'deeper' processing will have stronger facilitative effects than those requiring surface-level processing.

Using passages that described a number of psychological

principles, Watts and R. C. Anderson (1971) compared the effects of 'name' questions, 'repeated example' questions and 'application' questions. Name questions required the recall of the name of the psychologist associated with the principle. Repeated example questions involved the recognition of the examples that were provided in the material as instances of the principles. These questions were identical to the application questions except that the latter involved instances that were not stated in the passage.

Felker and Dapra (1975) examined 'verbatim' versus 'comprehension' questions, Andre and Sola (1976) and Andre (1981) compared the effects of 'verbatim' to those of 'paraphrased' items, and Friedman and Rickards (1981) "manipulate[d] depth of processing by inserting in text verbatim, paraphrase, or inference questions" (p. 428).

The common finding in these and other studies is that subsequent performance benefits more from higher-level (e.g., application) questions than it does from lower-order (e.g., recognition) items. In fact, the latter may even have inhibitory effects (Watts & Anderson, 1971). However, here again, opposite results have been found. For instance, in a series of seven experiments, Andre and his coworkers (Andre, Mueller, Womack, Smid & Tuttle, 1980) found no evidence for the superiority of 'application' questions over 'factual' questions. Similarly, in a study with children, Swensen and Kulhavy (1974) did not find a significant difference between 'paraphrase' and 'verbatim' questions, and Bing (1982) found 'rote' questions more helpful

than 'conceptual' questions. Also, despite their genuine efforts, R. C. Anderson and Biddle (1975) failed to show the superiority of 'paraphrased', as compared to 'verbatim' questions. This may be due to methodological problems (Andre & Sola, 1976). Alternatively, it may be the case that, contrary to what has been suggested by previous research (e.g., Sachs, 1976), memory for the surface features of text is not necessarily transient (see Hayes-Roth & Thorndyke, 1979; Walker, & Meyer, 1980). Furthermore, Carrier and Fautsch-Patridge (1981) have strongly criticized the idiosyncratic nature, and thus, the incomparability of the labels assigned to different types of questions across studies and have argued that this problem together with the ambiguity of experimental instructions to subjects can explain the inconsistencies found in the effects of higher-order questions.

More recently, there have been attempts to analyze the relationship between a text and questions generated relevant to it. These attempts are concerned with the semantics of discourse and since they are often based on discourse processing notions, they will be discussed, later, in that context.

Attempts to Account for the Effects of Adjunct Questions

Adjunct questions studies tend to be atheoretical. R. C. Anderson and Biddle (1975) have qualified this literature as being "infected with a mindless empiricism" and have argued that "we do not need another demonstration that adjunct questions 'work' [... we need to know ...] why they work and under what conditions" (p. 108) (for similar remarks, see Andre, 1979, p.

302; McConkie, 1977, pp. 38-39).

There has been, however, a number of attempts -for the most part post hoc- to provide an explanation for the effects, especially the indirect effects, of adjunct questions. Rothkopf's (1965, 1966; Rothkopf & Bisbicos, 1967) 'mathemagenic hypothesis' is the best known. The term 'mathemagenic' characterizes behaviors that give birth to learning. The idea is an attempt to explain the indirect effects of adjunct questions and is based on the behavioristic notion of shaping: "Tests are reinforcement-like events for certain desirable mathemagenic responses" (Rothkopf, 1965, pp. 216-217). Successful answering of adjunct questions reinforces the students' mathemagenic behaviors and thus improves their posttest performance. Failure to answer the adjunct questions correctly will lead to the extinction of mathemagenic behaviors. Since the probability of succeeding is higher when the items appear after the relevant passage, such items give better results than those placed before the text (Rothkopf & Bisbicos, 1967).

Numerous studies (e.g., Bruning, 1968; Felker & Dapra, 1975; Frase, 1967; McGaw & Grotelueschen, 1972) have interpreted their results as support for the mathemagenic hypothesis. As pointed out by Anderson and Biddle (1975) and Rickards (1979), over the years the strong behavioristic flavor of the mathemagenic hypothesis has been replaced with a cognitive language. Rothkopf's (1972, 1976) more recent writings emphasize mathemagenic 'processing activities' rather than behaviors and the concern for the effective, as opposed to the nominal, stimuli

is now expressed in terms of the internal representation of the text.

Following a number of difficulties for the mathemagenic hypothesis (see e.g., Frase, 1968b) and due to the shift of emphasis in experimental psychology toward a cognitive perspective, a number of more process-oriented accounts of adjunct questions effects have been developed.

Frase (1967) has proposed two hypotheses, namely, the 'forward' and the 'backward' hypotheses. The former asserts that, in a forward manner, questions optimize mathemagenic behaviors on subsequent text. According to the backward hypothesis on the other hand, questions placed after passages lead to the mental review of what has been just read -hence, the better retention of the preceding material.

McGaw and Grotelueschen (1972), among many others, have provided evidence for both shaping (i.e., forward) behaviors and a backward, review process. Rickards (1979) has postulated four types of mental processing; namely, specific or general forward processes and specific or general backward processes. A specific process is limited to questioned material, while a general process is not. Rickards has shown how these four processes can account for the results of several adjunct questions studies.

R. C. Anderson and Biddle (1975) have argued that adjunct questions increase performance on repeated questions by facilitating mental review and further cognitive processing of the material. The levels-of-processing conception of the human memory system (Craik & Lockhart, 1972) has often been used to

explain the effects of adjunct questions. Roughly speaking, this framework can be viewed as an information-processing version of Bloom's (Bloom et al. 1956) taxonomy. Rather than considering all rehearsal and review processes to be useful, it emphasizes the quality of the processing. As Bruning's (1968) study has shown, merely restating the information, and thus providing the possibility of review, has lower facilitative effects than asking questions (for a related issue see Ellis, Konoske, Wulfeck II & Montague, 1982).

Levels-of-processing notions can be used in interpreting the results of several studies (e.g., Andre & Sola, 1976; Felker & Dapra, 1975; Watts & Anderson, 1971). Also concerned with depth of processing, Reder (1980) has emphasized the role of elaboration in learning from text and has argued that useful elaborations generated during initial question answering produce improved performance. She shows how the extra-elaboration notion can account for several sets of results.

## Methodological and Theoretical Issues Involved in Adjunct Questions Studies

In a review of various study aids, T. H. Anderson (1980) concludes that "the systematic use of good adjunct questions is the most effective one" (p. 500). It is true, however, that the volume of research in this area exceeds the number of reliable and adequate descriptions it has generated.

Adjunct questions studies have been criticized for their ecological invalidity (Rickards, 1980), and methodological shortcomings such as absence of appropriate control groups or

unreliability and vagueness of question classification schemes are seen as at least partial explanations for the inconsistencies in this area (Andre, 1979; Kormos, 1983; Carrier & Fautsch-Patridge, 1981). More important, however, are the theoretical limitations of adjunct questions studies (op. cit.). The area has tended to be almost exclusively functionalist: How posttest performance varies as a function of questioning and a number of variables related to it. Thus the studies are essentially product-oriented, and because of their minimal concern for the processes involved in question answering, they have not contributed much to our understanding of how people comprehend discourse and why some questioning strategies improve learning (Carrier et al. 1981; Reder, 1980).

Several investigators (T. H. Anderson, 1980; Andre, 1979; Carrier et al. 1981; McConkie, 1977) have persuasively argued for the centrality of students' intentions and their perception of the task demands and for the importance of explicit description of the requirements to the learners. Andre (1979) has proposed an information-processing model of learning from prose; in this model, questions are assumed to modify the learners' perception of the task and their cognitive representation of the material.

Furthermore, Carrier et al. (1981) have raised the "issue of the confounding of prose variables with question variables" (p. 376) and Reder (1980) has emphasized the necessity of a system of semantic representation of discourse for a model-based study of comprehension and question answering (see also R. C. Anderson, 1972; Kormos, 1983; McConkie, 1977; Rickards, 1979).

The study to be reported here has used one of the existing semantic representation systems (Frederiksen, 1975) and since the study used questions as a means of obtaining information about students' cognitive processes and not as potential pedagogical devices, the adjunct questions literature is not reviewed in further detail.

## THE THEORETICAL LITERATURE

With the shift of emphasis in experimental psychology from observable to internal phenomena and processes, cognitive operations such as comprehension and problem solving have become central research areas. While there exists a voluminous and rich literature on discourse comprehension, the process of question answering (Q/A) has not attracted comparable attention in cognitive psychology. As Lehnert (1978) has argued, this may be due to the fact that Q/A is too low-level and automatic for most of psychology (and too high-level for, say, neurophysiology). In fact, some of the cognitively oriented works on Q/A have been done by researchers in the field of artificial intelligence.

This section of the review is primarily concerned with the discourse comprehension literature. An overview of the research on the process of Q/A will be presented at the end of the section.

### Discourse Comprehension

#### Orientation and Basic Assumptions

The study of discourse processing by psychologists is primarily motivated by the assumption that discourse reflects cognitive structures and processes (Frederiksen, 1981). More

specifically, it is assumed that since memory is to a large extent semantic (Sachs, 1967), and since the meaning of a text is structured, memory structure may reflect discourse structure (Marshall & Glock, 1978; Stein & Glenn, 1979). Furthermore, this interest in memory structure has been accompanied with an interest in how knowledge structure is acquired. Here again, discourse comprehension is the natural place to begin. Current work in this area is, moreover, characterized by an emerging consensus that comprehension is more inferential than literal, more constructive than trace-abstractive, and more interactive than linear and sequential (Danks & Glucksberg, 1980; diSibio, 1982; Spiro, 1977, 1980a).

Evidence for the inferential and constructive nature of the human cognitive system goes back to the pioneering work of Bartlett (1932). For some of his recall experiments, Bartlett selected a story that was culturally distant from his subjects' world knowledge. He found that subjects 'remembered' highly distorted versions of the story. The reconstructions were in the direction of matching the story with one's prior knowledge, and showed a "tendency towards increasing conventionalisation of language" (p. 70). The reconstructions were even more marked in later recalls. Moreover, subjects were usually satisfied with their distorted versions.

There have been long debates over the validity of Bartlett's results (see diSibio, 1982 for a review). For instance it has been repeatedly argued (e.g., Kintsch, 1977; Meyer, 1977) that the distortions reported by Bartlett are to be attributed to the

unconventional nature of the story used and to the long delays between the presentation of a passage and its recall. In fact, empirical evidence (e.g., Kintsch & van Dijk, 1975; Thorndyke, 1977) suggests that accuracy of recall decreases when the textual input has an unusual structure and, in many studies (e.g., Kintsch, 1977; Kintsch, Kozminsky, Streby, McKoon & Keenan, 1975; Meyer, 1977) at least non-delayed recall has been found to be much more accurate than Bartlett reports.

However, Spiro (1980b) has persuasively argued that the relative contribution of prior knowledge is affected by the characteristics of the reading material, the purposes of reading, and individual differences. He has attributed the accuracy of recall found in some studies to the artificial context of memory experiments, and has argued that, typically, experimental as well as school instructions and materials "minimize interaction with and assimilation to prior knowledge " (1980a, p. 255; see also Spiro, 1977 and diSibio, 1982). Moreover the conditions under which inaccuracy occurs, rather than inaccuracy per se, are important (diSibio, 1982; Spiro, 1977). Reder (1980) contends that what Bartlett's subjects did is typical of discourse processing.

In fact, Kintsch (1977) explicitly states that in his study "subjects were admonished to be accurate" (p. 53). It is, therefore, not surprising to find "quite accurate" recalls. The issue seems to be one of competence-performance. Apparently, people can be quite accurate, but, as several social psychologists have argued (e.g., Greene, 1976; Hamilton, 1981),

people tend to be concerned with coherence and meaningfulness more than accuracy.

However, what has been controversial is the extent to which Bartlett's results are to be accepted (see diSibio, 1982 and Spiro, 1977 for discussions of the constructive/reconstructive issue). That preexisting cognitive structures interact with the input and affect memory processes seems fairly uncontroversial and despite the lack of systematicity that characterizes Bartlett's work, his study is frequently cited in current literature on memory and comprehension.

Yet, it should be noted that the willingness to study discourse comprehension with a constructive orientation is quite recent (Danks & Glucksberg, 1980; Meyer, 1977; Reder, 1980; Spiro, 1980a; Thorndyke & Yekovich, 1980). Bartlett's structuralist orientation was ignored by the dominant empirical-associationistic conceptions of the human mind (Kintsch, 1978; diSibio, 1982; Spiro, 1980a). For years, verbal learning psychologists restricted their material to nonsense syllables, isolated words, and unrelated sentences, that is, more easily controlled stimuli, free of previously learned associations (Danks & Glucksberg, 1980; Reder, 1980; diSibio, 1982). These kinds of material hardly lend themselves to deep, semantic processing (Craik & Lockhart, 1972) and to a constructive, inferential activity on the part of the learner (Lachman & Lachman, 1979).

Gradually, however, with increasing demands for generalizability of laboratory findings to educationally

meaningful tasks, with developments in computer science that allowed simulation of discourse processing, and with the development of case and text grammars in linguistics, the study of discourse comprehension came under the focus of psychologists (Reder, 1980). Evidence against the formalist position accumulated (Morgan & Green, 1980). Several experiments demonstrated that people construct cognitive representations that encompass information not explicitly encoded in the presented discourse (Rumelhart, 1980; Spiro, 1980a). A number of experiments conducted by Bransford and his coworkers (e.g., Bransford & Franks, 1971; Bransford, Barclay & Franks, 1972) have been influential. In one study (Bransford et al. 1972), college students received sentences like this: (1) "Three turtles rested beside a floating log, and a fish swam beneath them" or (2) "Three turtles rested on a floating log, and a fish swam beneath them". These two sentences have identical deep structures; yet, the semantic situations suggested differ. Subjects were then given recognition tests of sentences such as (3) "Three turtles rested (beside/on) a floating log and a fish swam beneath it". It was hypothesized that answers to the recognition test would vary depending on the initial sentences received. As predicted, and contrary to what would be expected by a traditional interpretive theory, it was found that subjects who first heard sentence (1) tended to reject sentence (3) on the recognition test. In contrast, subjects hearing sentence (2) were likely to think they had heard sentence (3) before. Thus, "recognition was shown to be primarily a function of the complete semantic descriptions

constructed rather than a function of just that information specified by the linguistic input strings" (p. 205). It was concluded that it is people and not sentences that carry meaning.

As Spiro (1980a) has pointed out, this study served as a prototype for numerous other demonstrations of the false recognition of new information that is compatible with the old information. Despite the challenges that this finding has been faced with (e.g., Hayes-Roth & Thorndyke, 1979; Katz, Atkeson & Lee, 1974; Katz & Gruenewald, 1974; see Walker & Meyer, 1980 for a discussion), the belief that the human memory system is inferential and constructive has become widespread. In fact, as suggested in an excellent article by Magoon (1977), presently, several disciplines -including cognitive psychology, social psychology, sociology, anthropology, and educational research- show a tendency towards a constructivist approach, hence, conceptualizing human subjects as purposive, meaning-seeking, knowing beings exhibiting rule-governed constructive behavior. Methodologically, this has led to a preference for 'thick descriptions' (Geertz, 1973) such as free recall protocols in discourse processing research.

As mentioned earlier, the interactive nature of language comprehension is now widely accepted (Danks & Glucksberg, 1980). It is believed that various sources of information including the comprehender, the author, the discourse and the context (e.g., task demands) continuously interact during language processing (Frederiksen, 1981; Tierney & Mosenthal, 1980) and that, across linguistic levels, processing is interactive (Adams & Collins,

1979). The notions of 'bottom-up' and 'top-down' processing have, thus, become commonplace (Adams & Collins, 1979; Bobrow & Norman, 1975).

In bottom-up conceptions of language processing, the major role is attributed to the textual information, and lower-level processes are presumed to occur prior to higher-level processes (Frederiksen, 1979). In top-down models, high-level inferential operations are thought to control the comprehension process (op.cit.), and the comprehender's existing knowledge structures are viewed as the major contributors of the process (de Beaugrande, 1981). The bottom-up/top-down dichotomy has also been called 'data-driven' and 'conceptually driven' (Bobrow & Norman, 1975). Data-driven processes go from part to whole, whereas conceptually driven processes go from whole to part (Rumelhart, 1980). The former searches structures in which to embed the data; the latter fits the input to expectations and predictions (Bobrow & Norman, 1975). The terms 'text-based' and 'knowledge-based' (or 'schema-based' or 'frame-based') refer essentially to the same phenomena (Frederiksen, 1977b; Spiro, 1980a). The study to be reported here used two types of questions: 'Text-based questions' and 'frame-based questions'. The distinction is analogous to that of bottom-up and top-down processing. More details will be provided in the Method section.

Most discourse processing models assume, and ample evidence suggests, that both top-down and bottom-up processing occur in comprehension. In a study of eight-year olds' inferences in answering questions about narrative discourse, Nicholson and

Imlach (1981) have observed both types, with bottom-up processing showing some priority over top-down processing. The nature of the interaction between top-down and bottom-up processing, however, is not well understood and the literature about it remains, for the most part, quite sketchy. Attempts to design artificial information-processing systems have clearly demonstrated the complexity of the interaction (Bobrow & Norman, 1975) and there is a controversy over the extent to which each type of process is involved in discourse comprehension (Frederiksen, 1979).

Cognitive psychologists differ from the experimental psychologists who preceded them in various ways. Central to these differences is the importance assigned by cognitive psychologists to two concepts: (1) The representation of knowledge and (2) cognitive processes. From a constructivist standpoint, knowledge is assumed to be represented in organized holistic structures. The organization is highly rule-governed rather than, say, associationistic (Frederiksen, 1977a). The representation/process emphasis can, thus, be regarded as a concern for structures and processes.

The following review is organized according to the structure/process distinction. However, since separating structures and processes is often unrealistic -processes act upon structures and structures include processes (see Rumelhart & Ortony, 1977, p. 127; Smith, 1977)-, overlaps and arbitrary separations are inevitable.

## Representational Discourse and Knowledge Structures

There exists a variety of theories about the form of cognitive representation of meaning. Some are analogical in orientation, many are propositional and schematic. The purpose of this section is not to review theories of knowledge representation (Norman, 1976, Chapter 8 provides a good overview; McConkie, 1977, pp. 14-15 is a quick reference; see also Chafe, 1977; Frederiksen, 1977a and Smith, 1977). The purpose is the selective review of the main theories of representation, in memory and in text, that are relevant to discourse processing issues.

It is important to mention that the section is not limited to 'purely' conceptual structures. Although many of the propositional and linguistic representation theories may not claim any analogy between their system and psychological structures, these theories are still primarily structural ones and will be discussed in this section.

At the 'lowest'[1] level, are the textual features of discourse that affect comprehension. Next[1], are the propositional structures, and finally[1], are the highly organized conceptual structures that comprehenders have or construct and that writers and speakers instantiate in their discourse.

### Textual features of discourse that affect comprehension

As the unit of analysis in linguistics and psycholinguistics has, to a large extent, shifted from sentence to discourse (Danks

---

1 Obviously, no psychological linearity is implied here.

& Glucksberg, 1980) a bulky literature on the linguistic structure of text has been generated, and numerous experiments trying to show the cognitive effects of textual variations have been conducted.

While it is generally held that the property that differentiates discourse from a random sequence of sentences is 'coherence' (Frederiksen, 1977b; see de Beaugrande 1980, 1981 for 'principles of textuality'), no agreed-upon definition of discourse exists in the literature, partly because the term 'coherence' makes sense primarily in relation to a context and to a reader or listener. However, discourse is generally seen as structured units that display some organizational structure (Meyer, 1977) with various cohesive devices that link constituents together (Danks & Glucksberg, 1980).

Empirical research has demonstrated that the processes comprehenders use and the knowledge they acquire from discourse are affected by such surface features as cohesive items and topicalization patterns of text. For instance, Garrod and Sanford (1977) found that, where the integration of two anaphorically related items is necessary, reading time and, presumably therefore, comprehension processes vary as a function of the 'semantic distance' between the two items.

In fact, Halliday and Hasan (1976) have proposed the notion of 'cohesion' as a factor contributing to discourse coherence. Cohesive ties establish a text's continuity and occur "where the INTERPRETATION of some element in the discourse is dependent on that of another. The one PRESUPPOSES the other" (p. 4). Halliday

and Hasan propose a classification of cohesive ties in four main groups: Reference, substitution and ellipsis, conjunction, and lexical cohesion; each category is further subdivided. Moreover, two types of reference are distinguished: 'Exophoric' and 'endophoric'. Exophoric reference is situationally-determined, i.e., it signals that reference must be made to outside the text. Endophoric reference, on the other hand, is textual and is assumed to refer either to preceding text (anaphora), or to following text (cataphora).

Tierney and Mosenthal (1980) have pointed out that cohesive ties have a mortar-like quality and that they play a major role in determining the affective power of a text. Morgan and Sellner (1980), however, have criticized Halliday and Hasan for "the mistake [they...] have made [...] -taking certain aspects of linguistic form as cause, rather than effect, of coherence" (pp. 179-180). Halliday and Hasan, however, do mention that cohesion is not the only condition for the creation of text. 'Register' is another necessary condition. While cohesion makes a text coherent with respect to itself, register makes it coherent with respect to the context.

In addition to cohesive ties, the thematic organization of discourse has been found to signal or require certain processes and, thus, to affect comprehension. For instance, Clements (1979) has provided strong evidence that 'staging', a notion borrowed from Grimes (1975), affects what we remember from discourse. In this nicely designed and reported study, staging was defined as "a dimension of prose structure which identifies the relative

prominence given to various segments of prose discourse" (Clements, 1979, p. 287). The notion is based on the idea that the combination of the topic-comment and the given-new distinctions gives a hierarchical organization to discourse. Through the identification of the topic of a text segment, its novelty, and its relationship with an earlier segment, the hierarchical level that was 'assigned' by the writer/speaker to that segment can be determined.

Clements hypothesized that, in the absence of some higher level of control (e.g., strong feelings about non-emphasized points of the text), staging affects the memorability of information. He found that a given piece of information was better recalled if it staged high than if it staged low. The psychological status of staging was also demonstrated by Marshall and Glock (1978) who found that, independent from variations in the semantic structure of discourse, staging affects recall[2].

Closely related to Grimes's (1975) notion of staging is Meyer's (1977a, 1977b; Meyer, Haring, Brandt & Walker, 1980) technique of discourse analysis. This technique yields hierarchical tree structures that show the superordinate/subordinate relationships between the ideas of a passage. Despite some confusion that seems to exist in this framework between the conceptual importance of a chunk of information and its organizational 'height', results reliably indicate that recall is significantly affected by the hierarchical level of information: Top levels are recalled and

---

2 There were group differences, however.

retained better than lower levels.

Kintsch and his colleagues (Kintsch, 1978; Kintsch, Kozminsky, Streby, McKoon & Keenan, 1975; Kintsch & van Dijk, 1975) have also found that 'superordinate' propositions are better recalled than 'subordinate' ones. Kintsch (1978) has attributed this finding to the greater degree of implicit repetitions that superordinate propositions receive in a text, and has related it to the notion of the given-new strategy of text comprehension formulated by Clark and Haviland (1977; Clark, 1977) in the study of conversational conventions.

Meyer (1977a, 1977b) has suggested that a theory of writing test questions can be derived from her framework. She argues that, unlike Bormuth's (1970) and Anderson's (1972), her approach is useful in deciding which information to query (1977a). However, as Resnick (in the 'open discussion' chapter following Meyer, 1977a) has pointed out, what psychometricians need is knowledge about memory processes; such knowledge is not provided by Meyer's framework.

Semantic structure of discourse

With the accumulation of evidence (e.g., Sachs, 1967) that memory is to a large extent semantic, cognitive psychologists and researchers in artificial intelligence have tried to find ways of representing the semantic structure of discourse. The psychologist's motivation to do so is partially derived from the assumption that a comparison of the semantic representation of a passage (input) and that of a recall protocol (output) allows one to infer the cognitive processes that occur between the stimulus

and the response, and to infer the knowledge acquired.

Helped by developments in linguistics -e.g., Fillmore's (1968) case grammar emphasizing the semantic relations among words-, a number of models have been proposed for the conceptual representation of discourse (e.g., Frederiksen, 1975; Kintsch, 1974; Norman, Rumelhart & the LNR Research Group, 1975; Schank, 1972). Among these, the systems developed by Frederiksen and Kintsch have been used most in psychological studies of discourse processing. Both systems consider the proposition as the unit of meaning, both use a propositional, rather than a network, notation, both yield a hierarchical structure that allows for multiple embedding of propositions, and both have the potential for representating various content areas and discourse genres (see de Beaugrande, 1981, and Tierney & Mosenthal, 1980 for comparisons of the two models).

In Kintsch's (1974) system, a connected, ordered list of propositions forms the 'text base'. A proposition contains a predicator and a sequence of arguments. Arguments are word concepts (e.g., nouns). A predicator, on the other hand, is a relational term (e.g., verbs) that connects arguments. Thus, a relation and a set of arguments represent a unit of idea. The comprehender's task is regarded as one of inferring the 'text base structure' of the speaker or the writer, and constructing a representational structure of it (Kintsch, 1978). Several studies (e.g., Kintsch, 1974; Kintsch, Kozminsky, Streby, McKoon & Keenan, 1975) have demonstrated that what is important in reading comprehension is the propositional content of discourse rather

than surface structure.

For a text base to be coherent, a number of conditions must be met. One of these conditions is referential coherence (Kintsch & van Dijk, 1978). This is done through a repetition rule: Each proposition must share an argument with at least one other proposition. Propositions containing repeated arguments are 'subordinated' to the proposition where the argument originally appeared; 'superordinate' propostions contain 'subordinate' propositions as their arguments (Kintsch, 1974, 1978). While not being considered a sufficient or even a necessary condition for coherence, the repetition rule is viewed as having substantial predictive value for recall, and as mentioned earlier in the context of Meyer's theory, Kintsch reports empirical results in support of the higher retention of superordinate propositions (Kintsch, 1978; Kintsch et al. 1975; Kintsch & van Dijk, 1975).

Kintsch (1974) assumes an analogy between his representational system and the representation of meaning in memory. Methodologically, the system is simple to use and it has proved useful in numerous empirical studies.

Defining comprehension as the process of building a model for discourse and for the conceptual structures that underlie it while satisfying text-based, knowledge-based, and context-based constraints, Frederiksen (1975, 1981) has developed a semantic representational system that is, in its basic principles, similar to Kintsch's (1974). Perhaps their main difference is in their degree of elaboration: Kintsch's technique is sketchy and loose-hence its easiness; Frekeriksen's (1975) is elaborate and

specific -hence its richness.

The system is based upon concept-relation-concept triples. Two classes of concepts have been distinguished: 'Objects' and 'actions'. Objects are things occupying space. Actions are things occupying a position or interval of time and involving change. Each category is further divided into subclasses.

Concepts are connected to each other by a network of defined and labelled binary relations. There are different types of relations: 'Case' relations specify an action; 'identifying' relations, on the other hand, distinguish an object or an action from other objects or actions. When a concept-relation-concept triple involves objects and identifying relations, it defines a 'state'. In contrast, an 'event' involves actions rather than objects but can have both types of relation. In propositional notation, each event or state forms one proposition (Frederiksen, 1981).

In addition to representing the components of a proposition, the system can handle larger chunks by representing the 'relative' and 'dependency' relations across propositions. The model is therefore hierarchical in nature with the concepts at the lowest level and the interpropositional relations at the highest level (Frederiksen, 1977a).

The framework (Frederiksen, 1975) is actually much more detailed than is suggested by the above overview. The study to be reported here used a simplified and updated version of the initial model (see Frederiksen, 1981).

It must be mentioned that the system is intended as a theory

of memory representation at the propositional level. Marshall and Glock (1978) have found that recall varies as a function of manipulations in the semantic structure of discourse, as represented by Frederiksen's (1975) system. These researchers indicate that if we accept the assumption that discourse structure reflects memory structure, a demonstration of the descriptive adequacy of the model at the discourse level equally supports it as a model of memory.

Frederiksen's system has been used in several studies which have shown its descriptive adequacy. Relevant to question answering issues, a study by Kormos (1983) must be mentioned. In this study, children read a story, answered questions relevant to it, and retold it in their own words. The questions were generated on the basis of Frederiksen's propositional (1975) and frame (1982, in prep.) analysis systems (the latter system deals with more abstract conceptual structures and will be discussed later). There were three types of question: 'Factual' questions interrogated case information that was explicitly encoded in the text; 'connective' questions required subjects to infer implied interpropositional connections; finally, 'framing' questions were about the underlying global structures (e.g., questions about motives and plans of the story characters). Recall protocols were found to vary depending on the type of question and reading level. While framing questions were not facilitative for these grade-three children, aptitude-treatment interactions were found in the case of connective questions. Moreover, asking questions seemed to focus recall toward propositions relevant to the type

of question.

This study has at least two advantages over many of those reviewed in the Adjunct Questions section. First, we know exactly what the labels assigned to different question types refer to; using the same propositional and frame analysis systems we can replicate the study. Second, by including different reading levels, different passages, and different response modes (recall versus inference), the study has attempted to provide data that are less and less dependent on a given experimental situation. In fact, high-level interactions among these variables were found, thus, showing the complexity of the effects of questions.

Relevant to questioning issues and Frederiksen's (1975) model, there is another work that is to be noted: Lucas and McConkie (1980) have proposed a way of analyzing questions. In this framework, a question is defined in terms of (1) the propositions that pertain to answering it and (2) its relation to the text. The former is done on the basis of Frederiksen's system. The latter, on the other hand, is defined in terms of a set of descriptors (e.g., 'stated', 'implied', etc.). The framework is thus descriptive rather than generative.

Global forms of conceptual representation

So far in the review of structure-based works, the focus has been on textual features and 'lower' levels of cognitive representation. In this section, the emphasis is upon more global forms of representation that deal with the structure of prior knowledge and expectations. Central to this knowledge-based approach is the notion of 'schema' or 'frame' that, under various

names, has been used extensively in both cognitive psychology and artificial intelligence. The notion is actually an old one and goes back at least to Bartlett (1932) and, of course, Jean Piaget used the notion throughout his work (see e.g. Rumelhart, 1980 for earlier references). Bartlett interpreted the results of his perception and recall experiments by the process of 'effort after meaning' which is "the attempt to connect something that is given with something other than itself" (p. 227). The latter 'something' refers to preexisting world knowledge which is assumed to be represented in highly structured holistic entities called 'schemata'.

More recently, discourse processing researchers have provided extensive data consistent with the notion of schema. A study by Dooling and Lachman (1971) is often referred to as a demonstration of the effects of schemata on language comprehension. In this study it was shown that very difficult, i.e., vague and metaphorical, passages can become comprehensible when readers are provided with a short title reflecting the unifying theme of the passage. Presumably therefore, once the appropriate schemata are suggested and activated, ambiguous texts become easy to understand (Rumelhart, 1980).

In another classic schema-demonstration experiment, R. C. Anderson and Ortony (1975) have shown that context and world knowledge interact to create holistic representations of discourse. In this study, subjects reading the sentence "televisions need expert repairman" were more effectively cued at recall by the word "appliance", while subjects reading the

sentence "televisions look nice in family rooms" preferred the cue "furniture". It is argued that language processing involves the dynamic construction of particularized and elaborated mental representations.

While each particular formulation of schema theory differs from the others, schemata and frames can be viewed as "data-structure[s] for representing [...] stereotyped situation[s], like [...] going to a child's birthday party" (Minsky, 1975, p. 212), "active processing elements which can be activated from higher level purposes and expectations, or from input data" (Bobrow & Norman, 1975, p. 132), "higher-level organizing principles [...that] unify concepts [...] under the constraint of typicality and normality" (van Dijk, 1977, p. 21), "data structures for representing the generic concepts stored in memory" (Rumelhart & Ortony, 1977, p. 101), "the structures of expectations " (Tannen, 1979, p. 138), "the building blocks of cognition " (Rumelhart, 1980, p. 33), or "cluster[s] of knowledge that describe [...] the typical properties of the concept [...they] represent" (Thorndyke & Yekovich, 1980, p. 23).

Schemata exist at all levels of abstraction and can hierarchically embed subschemata. They have 'slots' with variable names that, if associated with the elements of an input, 'instantiate' the schema. Since no instantiation matches the schema perfectly, evaluation of goodness of fit is a primary activity of schemata. Often, even if a schema matches the input, some of its variables are not instantiated and will therefore be assigned 'default values' (Adams & Collins, 1979; Minsky, 1975;

Rumelhart, 1980; Rumelhart & Ortony, 1977; Spiro, 1980a).

In schema-based approaches to discourse processing, comprehension is viewed as a kind of pattern recognition and is essentially regarded as a process of filling the slots of a fixed schema (Frederiksen, 1981; Kintsch, 1977; Rumelhart & Ortony, 1977; Winograd, 1977). It is analogous to the processes of hypothesis testing and theory verification (Rumelhart, 1980). Its constructive nature is to some extent due to the default assignment process (Spiro, 1980a).

The notion of schema is obviously a very powerful one; it makes a great deal of intuitive sense and it accounts for a large class of psychological findings of the type Cook and Campbell (1969) call "'stubborn facts' that 'speak of themselves'" (p. 24); moreover, it has been used not only in cognitive psychology and artificial intelligence but also in a variety of other disciplines, including social psychology, linguistics and sociolinguistics, sociology, anthropology and ethnography (see Tannen, 1979 for a discussion of the multidisciplinary nature of the notion). Despite the general acknowledgement that the notion is nebulous, ill constrained, and metaphoric (e.g., Bartlett, 1932; Hastie, 1981; Taylor & Crocker, 1981; Thorndyke & Yekovich, 1980; van Dijk, 1977), that it lacks predictive power and is unfalsifiable (e.g., Hastie 1983; Thorndyke & Yekovich, 1980), and that it is static (e.g., Collins, Brown & Larkin, 1980; Frederiksen, 1981; Tannen, 1979), few seem willing to relinquish its plausibility and its descriptive power.

Different types of schema have been contrasted in the

literature. For instance Frederiksen (1982) has distinguished 'content frames', 'context frames' and 'text frames'. Content frames represent knowledge of specific content about different situations. Context frames represent the structure of the task, and text frames involve knowledge about the structure of discourse at different levels. Winograd (1977, p. 75 ff.) has provided a similar classification and van Dijk (1977) has contrasted structures that deal with semantics with those concerned with organizational rules.

Several investigators have attempted to formally describe specific types of frames. Often -but not always (e.g., Frederiksen, 1981)- these models assume that comprehension is the mainly top-down process of fitting preexisting frames to stereotypical input. For instance, assuming that people possess schematized expectations about the typical structure of narratives in their culture, a number of theorists have proposed 'story grammars' or formal structural descriptions of the syntax of this type of discourse (Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, 1979; Thorndyke, 1977). Assuming that narratives are structured not only at the sentence level, but also at the suprasentential level (Rumelhart, 1975), story grammars attempt to describe "the legal combinations of abstract narrative elements" (Thorndyke, 1977, p. 78).

Story grammars have a transformational-generative-grammar flavor and consist of a hierarchy of basic elements such as setting, episode and resolution, and a partially ordered set of rewrite rules that map the story's 'deep structure' and its

surface structure. For instance in Rumelhart's model -from which the other grammars were derived-, a setting and a number of episodes are the main elements of the grammar. The former has slots for the time, the place, and the main characters of the story. Episodes, on the other hand, include events and reactions to it. Thus, the syntax of a well-formed story can be written in the form of a tree structure, the nodes of which represent the presumed slots of the schema which are instantiated with appropriate text propositions. With the objective of facilitating the "development of a story schema" in children, Sadow (1982) has proposed a procedure for deriving comprehension questions on the basis of Rumelhart's grammar.

That people are sensitive to the structure of stories, that there are developmental differences in knowledge of narrative structure, and that it is more difficult to understand stories that depart from a conventional structure have been demonstrated in several experiments (e.g., Kintsch, 1977; Kintsch & van Dijk, 1975; Mandler & Johnson, 1977; Stein & Glenn, 1979; Thorndyke, 1977). Moreover, Mandler and Johnson have applied their grammar to Bartlett's (1932) 'War of the Ghosts' story and have argued that the distortions reported by Bartlett are due to his story's numerous omissions and violations of the 'ideal' structure.

While story grammars are quite popular, and because of their simplicity they can be used efficiently as research tools (Tierney & Mosenthal, 1980), several criticisms have been addressed to them. One of the major problems of story grammars is shared by schema theory in general: The emphasis on preconceived

schemata with invariant slots and the over-reliance on top-down processing in terms of well-formedness inevitably restrict the applicability and the psychological reality of these models (de Beaugrande & Miller, 1980; Frederiksen, 1981). This issue will be elaborated in the Process section of this review.

The psychological reality of story grammars has been questioned. Bruce (1980b) and Reder (1980) indicate that these grammars are not process models. Frederiksen (1981) doubts that they could be psychologically more valid than derivational grammars at the sentence level and Black and Bower (1980) report recall data that do not reflect the hierarchical structure that story grammars assign to narratives. The limited applicability of story grammars is also often pointed out. Reder (1980) indicates that she was not very successful in applying Rumelhart's (1975) grammars to even simple stories. She attributes this problem to the fact that the grammar is limited to one-setting narratives. Similarly, Tierney and Mosenthal (1980) point out that story grammars have difficulty handling stories that include more than one protagonist. The importance of the interactions among several protagonists' plans in producing episodic structure has been emphasized by Bruce (1980a, 1980b) who points out the failure of story grammars to capture this aspect.

Bruce (1980b) also indicates that since the criteria for parsing stories are not specified, applying the grammars involves interpretation and agreement problems (see also Morgan & Sellner, 1980). Frederiksen (1981) has pointed out that when a particular structure is not specified in advance it is very difficult to

generate stories on the basis of story grammars. Black and Bower (1980) have not only criticized story grammars as formal systems, but have also indicated that "there are [...] many stories that the story grammars will not generate [...and] they accept non-stories as stories" (p. 231). The authors add however that the distinction between stories and non-stories is very vague.

Story grammars are limited to folktale-like narratives. Thorndyke and Yekovich (1980) report data suggesting that distinct frames may exist for different types of discourse. Frederiksen (1982, in prep.; C. H. Frederiksen & J. D. Frederiksen, 1981; Bracewell, C. H. Frederiksen & J. D. Frederiksen, 1982) has described a variety of frame, e.g., conversational, narrative, procedural, descriptive, explanatory and problem frame, for different types of discourse. Building upon his model of propositional representation, Frederiksen defines a frame in terms of conceptual elements and frame relations that connect them into higher-order units. For example, a narrative frame is composed of such elements as events linked by relations such as temporal and causal, to form units such as episodes. Or, a problem frame involves elements such as a problem state, a plan, acts, blocks and revisions connected by hierarchical relations to form, for instance, procedural hierarchies (Frederiksen & Frederiksen, 1981). There is a similarity between Frederiksen's conceptualization of a problem frame and Bruce's (1980a, 1980b) work on building representations for the plans and beliefs of story characters.

Frederiksen (1982, in prep.; Frederiksen, Frederiksen, &

Bracewell, in press) has devised procedures for the frame analysis of discourse. These procedures yield a network structure specifying frame elements and relations, and the propositions that instantiate them. Moreover, 'frame grammars', describing the rules for forming frames, have been developed. Ongoing research (op. cit.) is providing empirical support for the psychological status of the notion of frame.

It must be mentioned that Frederiksen's notion of frame differs from that of the 'typical' schema theorist in that central to it are the processes of frame construction rather than mere frame instantiation. This issue will be discussed in the Process section. Moreover, since this study used Frederiksen's notion of frame, further references to it are made in other sections of this work.

While narrative schemata of story grammarians and Frederiksen's frames are concerned with representation of knowledge about discourse structure, a number of theorists have attempted to formalize the organization of world knowledge, e.g., knowledge about interpersonal and social interactions (e.g., Bruce 1980a, 1980b). In this area, Schank and Abelson's (1977) notion of 'script' has been very influential. A script is a context-specific schema involving the frequently encountered and expected sequence of events and actions in a well-specified mundane situation. It is a scenario for recurrent patterns of social life within a culture (Abelson, 1976). For instance, a restaurant script, a frequently used example, includes the sequence of such events as entering, being seated, ordering,

eating, asking for the bill, paying it, and leaving. Typically, stories involving a restaurant episode omit many aspects of the complete script. Yet, since people presumably have the script in their heads, they can predict the missing elements.

Script theory has been developed with the purpose of building language understanding computer programs, and it has an extremely top-down conception of discourse processing. However, its psychological validity has been supported in studies showing the facilitative role of presumed script-instantiation in story understanding (e.g., R. C. Anderson, Spiro & M. C. Anderson, 1978; den Uyl & van Oostendorp, 1980).

Like many other schema theories, script theory is limited to areas of knowledge that have already been stored. As Tannen (1979) has suggested, a static view of schemata is characteristic of works that come out of artificial intelligence laboratories. Reder (1980) has pointed out some of the difficulties of script theory and has emphasized the elaborative and reconstructive processes of language comprehension and their idiosyncratic nature, aspects of discourse processing that are properly human.

In the context of questioning issues involved in reading instruction, Pearson and Johnson (1978) have loosely used the notion of script and have distinguished three types of question-answer relations, on the basis of the information source necessary to answer questions. 'Textually explicit' questions can be answered on the basis of the text alone. To answer 'textually implicit' questions, on the other hand, the reader must perform inferential operations on the information that is explicitly

encoded in the text. Finally, 'scriptally implicit' questions require the reader to read 'beyond the lines' by using his or her script, i.e., according to Pearson and Johnson, his or her prior experience.

This question-classification system was used in a study by Johnston (1982) who found that textually explicit questions were easier than textually implicit questions which were easier than scriptally implicit questions. Moreover, question type interacted with the centrality of the queried information and whether or not the reading material was available during question answering. Johnston argues that to assess comprehension, scriptal questions are best if integration to prior knowledge is central in our definition of comprehension. If, however, we view comprehension as the process of building a coherent model of the text, then we should ask textual questions that interrogate the central aspects of the text.

The end of this section on global conceptual representation seems like a good place to mention the work of Kintsch and van Dijk ( Kintsch, 1977, 1978; Kintsch & van Dijk, 1975, 1978; van Dijk, 1977) on 'macrostructures'. Macrostructure theory argues that, beside the local microlevel which consists of a linear propositional structure, a global macrolevel of description is needed to characterize discourse. Comprehension is essentially viewed as a process of assigning macrostructures to the propositional sequence (Kintsch & van Dijk, 1978). Macrostructure is the topic or the gist of discourse, it is its overall organization, and is expected to be recalled better than the

microstructure. Summaries in particular are expected to directly reflect macrostructures. Moreover, since macrostructures are viewed as necessary to comprehension, they are assumed to be constructed during reading rather than at time of recall. Kintsch and van Dijk (1975) report data that are consistent with these predictions.

Morgan and Sellner (1980) have criticized the notion of macrostructure for containing less information than the text whereas, according to Morgan and Sellner, the meaning of the whole text is more than the meanings of the parts. It must be mentioned however that Kintsch and van Dijk (1978) do allow macrostructures to contain information not represented in the original text. Moreover, one may opt for a different conception of text, for instance, de Beaugrande's (1981): "A text is often both more and less than the sum of its parts [...It is] an actual system" (p. 295).

Morgan and Sellner (1980) also criticize the vagueness of Kintsch and van Dijk's notions and especially their frequent reduction of content problems to linguistic problems. Earlier in this review, the lack of specification of Kintsch's work was pointed out. It seems that the sketchy nature of this work has lead to vagueness and even inconsistencies across articles. It has been our experience that psychologists expert in the area of discourse processing can differ markedly in their interpretation of the notion of macrostructure.

If macrostructure is the topic of discourse then the notion is primarily concerned with discourse organization and it should

have been reviewed together with Clements's (1979) and Meyer's[3] (1977a, 1977b) works. However, some of Kintsch and van Dijk's writings on story macrostructures seem just like another story schema theory (see e.g., Kintsch, 1977; Kintsch & van Dijk, 1975). In other writings, an attempt has been made to differentiate narrative frames and macrostructures (Kintsch, 1978; Kintsch & van Dijk, 1978; van Dijk, 1977). However, the distinction which seems to be one of syntax/semantics is neither very clear, nor consistently maintained.

Comprehension Processes

Despite the above promising title, it must be mentioned that relatively little is known about the process side of comprehension. For one thing, while we can ask people to read, recall, and answer questions, we cannot look at what is going on inside their heads. All we can do is try to infer processes from other sources of information.

Psychologists have used various techniques for studying cognition. Many of these techniques are product-oriented, e.g., free recall and question answering tasks, and some are process measures, e.g., reaction time, eye movements and on-line tasks (see Johnston, 1981). Process-oriented techniques, however, have not been used frequently in the area of discourse comprehension. Rather, they have been used primarily in problem solving research, psycholinguistic approaches to sentence comprehension,

---

3 In fact a similar confusion seems to exist with respect to Meyer's work. See, e.g., Rumelhart and Ortony (1977) for a schema-theoretic reading of Meyer's framework.

and those question-answering studies that aim at testing specific models of memory. The work of cognitively oriented discourse comprehension researchers on processes consists of primarily theoretical analyses and classifications of text-based inferences. Some attention has also been devoted to schema-related inferences and processes such as structure building and knowledge integration.

Within the constructive orientation of current work in discourse processing (diSibio, 1982; Spiro, 1980a), it is widely believed that comprehension is intrinsically inferential (Frederiksen, 1981; Kintsch, 1978). When a speaker or writer produces a text, much of what is intended is implicit and the comprehender has to bridge the discontinuities of discourse (de Beaugrande, 1981). The centrality of inference in language comprehension, even in young children, has been repeatedly shown (see e.g., Bransford et al. 1972; Frederiksen, 1972; Keenan & Kintsch, 1974; McKoon & Keenan, 1974; Nicholas & Trabasso, 1980), and it is generally held that frequently good and poor readers differ mainly in their inferencing processes, rather than in decoding and perceptual processing of the input (Frederiksen, 1979; Reder, 1980).

The fact that artificial intelligence researchers have to deal with inference (see e.g., Schank & Lebowitz, 1980) shows clearly that 'intelligent' comprehension necessarily involves inferential operations. Unlike some AI systems however (see Schank et al. 1980), good comprehenders know which inferences to draw and which not to draw (Warren, Nicholas & Trabasso, 1979).

Presumably they know this by recognizing the variety of constraints that they have to satisfy, namely, text-based, knowledge-based and contextual constraints, as well as validity constraints imposed by logical considerations (Frederiksen, 1981, in prep.).

Psychologists have been concerned with different levels of inference. The psycholinguistically oriented work on inference has primarily dealt with the pragmatics of spoken discourse (e.g., Clark, 1977; Clark & Haviland, 1977; Morgan & Green, 1980) and with text-based inferences, i.e., inferences that are signalled or required by linguistic features. For instance, the work of Garrod and Sanford (1977) which was referred to earlier is concerned with inferences required by anaphoric relations. These 'first-stage' inferences are a necessary prerequisite to most other inferences and their function is to complete the interpretation of a given sentence by resolving anaphora and ambiguities (C. H. Frederiksen, J. D. Frederiksen, Humphrey & Ottesen, 1978). Building a coherent representation of discourse also requires 'connective' inferences which link implicitly related propositions (Frederiksen, 1981; Frederiksen et al. 1978).

Inferences can be drawn to build a more or less elaborated representation of the text (op. cit). These inferences have been emphasized by van Dijk and Kintsch (Kintsch, 1977, 1978; Kintsch & van Dijk, 1975, 1978; van Dijk, 1977) who have postulated the notion of 'macrorule'. The argument is that inferential operators or macrorules act upon text base propositions to generate a

macrostructure which is a manageable unit for memory. A number of mapping rules which are used to generate macropropositions and to reduce the text to its gist have been identified. The macrorule of 'generalization' substitutes category names for category members. This rule would, for example, reduce a proposition about a cat and a dog to one about pets. The macrorule of 'deletion', on the other hand, eliminates propositions that are not needed in the interpretation of subsequent propositions. Finally, the macrorules of 'integration' and 'construction' eliminate propositions that represent the normal and expected antecedents, components, or consequents of a fact and find summary propositions that denote the global fact. For example, "Peter built walls. Peter built a roof..." becomes "Peter built a house". A number of constraints are involved in the operation of each rule. For instance, the application of the construction rule should lead to a proposition that is specific enough to be differentiated from other propositions. In the above example, "a human being did something" does not satisfy the constraint.

In a process model of comprehension and production, Kintsch and van Dijk (1978) assume that readers form a coherent text base by first checking out its referential coherence. Inference occurs when there is no argument overlap among all propositions. Macrorules are applied in cycles on the basis of increasingly stringent criteria of 'relevance'. Their application is under the control of schemata which are viewed as representing the comprehender's goals. Thus, it is assumed that one's goals determine which micropropositions will form the gist of the text.

A number of cognitively oriented researchers have been concerned with inferences that generate new propositions. Anybody who has seen recall protocols knows that generating new propositions is the rule rather than the exception of a free recall task and several investigators have attempted to classify these inferences (e.g., Frederiksen, 1979, 1981; Frederiksen et al. 1978; Nicholas & Trabasso, 1980; Rieger III, 1975; Warren, Nicholas & Trabasso, 1979).

Frederiksen's taxonomy of text-based inferences, for instance, attempts to specify the relationship between the comprehender's recall propositions and those of the presented material. The taxonomy is based on Frederiksen's (1975) propositional representation model and involves such categories as identifying operations and event specifying operations. More recently, Frederiksen (in prep.) has viewed inferences in terms of transformations that can operate upon various levels of data structures. The model is, therefore, no longer limited to text-based inferences; operations that act upon syntactic clauses, propositions and frames are conceptualized within the same framework.

Nicholas and Trabasso (1980; Warren, Nicholas & Trabasso, 1979) have also developed an inference taxonomy. Their taxonomy is specific to narrative discourse which they represent in terms of "event chains" depicting the logical relations that connect the sequence of episodes and actions in stories. Inference is regarded as serving a variety of functions, aiming at finding relations among story events. It is assumed that people make only

those inferences that are 'relevant' to the progress of the story.

The taxonomy of inferences includes five [4] main categories: Lexical, spatial, temporal, extrapolative, and evaluative. Lexical inferences permit one to choose between alternate meanings of words and resolve pronominal and nominal references. They answer 'Who?' and 'What?' questions. Spatio-temporal inferences encode the physical parameters of actions and determine the 'Where?' and 'When?' of stories. Extrapolative inferences address the 'How?' and 'Why?' of stories and are concerned with causes and consequences. Finally, evaluative inferences address the "So what?" of stories, and assess their significance, normality, and morality. The inference taxonomy is regarded as a potential basis for the systematic generation of questions to assess the inferential ability of children.

Unlike most story schemata, event chains are not presumed to be of psychological status (see Tierney & Mosenthal, 1980 for a comparison of story grammars and the event chain formalism). In this sense therefore, Trabasso and Nicholas' inferences are primarily text-based.

Schema theorists have done some work on the knowledge side of inferential processes. Because of their predictive nature, schemata are important sources of inferencing (Lehnert, 1978). Schemata resemble gestalt structures; hence, pattern-completion types of inferences, i.e., default assumptions, are typical of

---

4 The two cited articles differ in the number of categories and the labels assigned to them. They cover however the same processes.

top-down processing (see e.g., Adams & Collins, 1979; Kintsch, 1978; Rumelhart, 1980; Rumelhart & Ortony, 1977).

A central problem for schema theories however is that they view schemata as given (Frederiksen, 1981; Rumelhart & Ortony, 1977). Furthermore, that schemata undergo specification and modification in interacting with input (de Beaugrande & Miller, 1980) tends to be ignored by schema theories. It is not surprising to learn that the term 'schema' has been used by Kant (cited in, e.g., Thorndyke & Yekovich, 1980). Schema theories do tend to have an apriorist, fixist view of the human mind, and the schema-based/text-based distinction in the discourse processing literature reminds one of the age-old nature/nurture dichotomy. The distinction exists despite recurrent statements on the interactive nature of comprehension. One approach relies heavily upon internal resources; the other, upon environmental factors. Neither can account for the complex interactions among the two sets of variables.

The power of text-based theories lies in their ability to handle highly variable text formats (Frederiksen, 1981). These theories however fail to provide an account of the 'contextual appropriateness' of inferences (op. cit.). Moreover they are mainly concerned with representing the input and not enough with processing it (Reder, 1980), thus tending toward the 'meaning-is-in-the-text' fallacy (Spiro, 1980a).

Schema-based theories, on the other hand, have the power to handle the role of the comprehender's prior knowledge in inferencing. However, the link between schemata and textual

features and structures is not clear (Frederiksen, 1981). Moreover, these models cannot provide an adequate account of the comprehension of texts that have a high degree of structural variation or the processing of unfamiliar and non-stereotyped materials (op. cit.; Bracewell et al. 1982). Greeno (1974) has suggested that comprehension and learning are fundamentally similar in that they both are processes of constructing conceptual structures. The distinction between processing familiar material (in comprehension) and novel material (in learning) is, according to him, "quite soft and would not stand up to energetic attack" (p. 28). From such a perspective, an excessively schema-based view of comprehension cannot survive. A characterization of the processes through which schemata are generated becomes necessary.

A number of schema theorists themselves have addressed these issues. Even Bartlett (1932) contends that it is best to call schemata "active, developing patterns" (p. 201), and Spiro (1980a) asks the question: "Where do our knowledge structures come from in the first place?" (p. 260). Rumelhart and Ortony (1977) have attempted to describe processes of schema acquisition and modification. Thus the processes of 'schema specialization' and 'schema generalization' -both of which are important for learning- have been described. Rumelhart (1980) has also described the processes of 'accretion' or fact learning, 'tuning' or schema modification, and 'restructuring' or schema creation.

Yet, it is still argued that when the reader does "not have the appropriate schemata [...] he or she simply cannot understand

the concept being communicated" (Rumelhart, 1980, p. 48, emphasis added). As Spiro (1980a) points out, cognitive psychologists do not seem to be particularly concerned with learning. Paradoxically, the constructivist orientation has tended to be quite static.

Some of these problems have been addressed by Collins, Brown, and Larkin (1980) who conceive of comprehension as a model-building activity and propose that target structures are generative in nature. Using the terminology of the problem-solving research, the authors propose a 'progressive-refinement theory' that views text understanding as a process of constraint satisfaction. Collins et al. report retrospective 'thinking-aloud' protocols that reflect the progressive refinement of an initial model for the text which increasingly constrains the search for relevant data. Furthermore, the protocols show that readers evaluate a number of models while trying to make sense of a text.

In a similar line of argument, Frederiksen (1981) asserts that "an interactive theory of inference" is needed; a theory that "describes how a comprehender interacts with text to construct an interpretive frame that reflects both text properties and prior or contextual knowledge" (p. 304). Borrowing from ethnomethodological works on conversational inference that refer to the notion of frame as structures that are negotiated and mutually constructed by participants in a conversation, Frederiksen proposes a shift of emphasis in the conceptualization of knowledge structures (see also Tannen, 1979; Winograd, 1977).

In fact, in the study of natural conversation, the idea of fixed, a priori structures cannot survive because of the high degree of variability that typically exists in such contexts (see e.g., Dore, Geahart & Newman, 1978). 'Frame-construction theory' attempts to describe processes that are both text- and knowledge-based and assumes that frame-construction inferences[5] proceed according to rules analogous to the frame grammars that have been recently developed (Frederiksen, 1982, in prep., Frederiksen et al. in press; see also Bracewell et al. 1982). Frederiksen (1981) reports data consistent with the notion that comprehension is more than simple adherence to preexisting structures.

## The Cognitive Process of Question Answering

While question answering has interested the more socially oriented researchers, it has not been studied extensively from a cognitive perspective. This section attempts to outline some of the cognitively oriented works that have been done on the process of question answering (Q/A). Some of these works are computational in orientation and while they hardly constitute a cohesive research area, they tend to share two basic assumptions: (1) The processes involved in Q/A are different from those involved in comprehension and (2) Q/A is one of the best ways to demonstrate comprehension. The two assumptions may seem somewhat

---

5 In frame-construction theory, the term inference refers to processes that may not be called inference in other frameworks. This issue is to some extent theoretical; cognitive psychology has failed to differentiate some of the concepts that it uses most frequently (e.g., process, memory, recall, problem solving, comprehension, inference, etc.). The issue is however empirical too (see Kubes, 1982).

contradictory, but, more or less explicitly, they are often made.

One motivation for studying Q/A from a cognitive view is derived from the desire to build 'intelligent', interactive computer programs. Another motivation stems from the desire to create language 'understanding' computer systems: If Q/A is evidence for comprehension, then computerized 'comprehension' systems must be able to answer questions. Finally, the psychologist's motivation for studying Q/A comes from the desire to know more about the human cognitive system. Also, reaction time in Q/A has been extensively used as a means of testing specific models of memory.

On the basis of the age-old assumption that questioning can be an effective pedagogical device, a number of researchers have built computer-aided-instruction systems which are based on Q/A dialogues. For instance, Swets and Feurzeig (1965; Feurzeig, Munter, Swets, & Breen, 1964) have developed a computer teaching system that attempts to simulate tutorial conversations. More recently, Collins (1977) has formalized dialogues involving the Socratic tutoring strategy in a series of production rules. With the objective of modeling a tutor's role in a computer, Collins (Collins, Warnock & Passafiume, 1975) has analyzed the strategies that human tutors use and has synthesized, in a computer program, the results of his analysis. Collins (Collins, Warnock, Aiello & Miller, 1975) claims that his 'computerized Socrates' can help students "to reason in a generative way from incomplete knowledge" (p. 409). While the pedagogical value of Socratic dialogues has often been questioned (see e.g., Parlebas, 1980),

whether Collins' claim is reasonable or not remains an empirical question (Resnick, 1977). In any case, since they are primarily concerned with asking questions, the computer simulations of tutorial dialogues do not tell us much about the process of question answering.

On the other hand, cognitive scientists who work on natural language 'understanding' computer programs have been concerned with question answering per se (see e.g., J. R. Anderson & Bower, 1974; J. R. Anderson, Kline & Lewis, 1977; Hunt, 1975; Kaplan, 1981; Lehnert[6] 1978, 1981). Lehnert has developed a Q/A program which together with two story-'understanding' systems reads stories and answers questions about what was read. These systems have been developed at the Yale Artificial Intelligence Project and are based on Schank's (1972) system of semantic representation ('conceptual dependency') and on Schank and Abelson's (1977) formalization of 'scripts' and 'plans'.

As reflected in her program, Lehnert views Q/A as a three-stage process: 'Understanding' the question (parsing), finding an answer (retrieval), and translating the answer into language (generation). In order to 'understand' the question, the program must first categorize it. For instance, when a why-question is asked, does the interrogator want an answer about the cause of something or about its goal? Conceptual categorization of questions aims at determining exactly the type of answer the questioner wants. This is however not sufficient for understanding a question. Often, it is also necessary to infer

---

6 Lehnert (1978, pp. 263-273) reviews some of the best known computational query systems.

the intent of the questioner. Context becomes important in this phase. In many contexts a correct answer is not a good one, e.g., Q: Would you like to dance? A: Sure. You know anyone who wants to? (Lehnert, 1978, p. 5). Often, the answerer also needs to make assumptions about the interrogator's knowledge. This issue will be elaborated in the context of Norman's (1972) work on Q/A.

Once the question has been 'understood', an answer must be found. The retrieval process decides how much of an answer is needed. This content specification process depends on socio-psychological[7] variables such as the 'attitudinal' mode of the answerer. Finally, retrieval heuristics do the memory search and must select the best answer, i.e., answers that "convey the most relevant information in the most efficient way" (Lehnert, 1978, p. 188). This depends to a large extent on the assessment of the questioner's knowledge.

Lehnert's process model of Q/A is much more elaborate and explicit than what appears from the above outline, and her book (Lehnert, 1978) does an excellent job in presenting the model. It remains, however, that the model is a computational one and involves the fixity that characterizes the current state of computer science. Lehnert's answers tend to be 'found' or 'selected' rather than constructed. As Chafe (1977) has suggested however, human language generation is a creative process (see also Ebbesen, 1981), and Reder (1979) has provided experimental support for the notion that human question answering involves

---

7 For a comprehensive theory of Q/A, therefore, a theory of conversation is needed.

considerable manipulation of the retrieved information[8].

Norman (1972; Lindsay & Norman, 1972) has made a strong case for the view that question answering is more than simple retrieval from a static information store. Consider the question "What was Beethoven's telephone number?". When asked a question such as this, people do not start searching their memory. Rather, they reject the question as illegitimate. Using examples such as this, Norman argues that Q/A is preceded by considerable preprocessing, interacts extensively with one's world knowledge, is inferential, and proceeds by stages. First, there seems to be a preliminary analysis of the question itself. If the analysis leads to the decision of attempting a recall, search strategies are set up. Considerable processing occurs during retrieval too; for instance, the stored information that is relevant to the question is evaluated before being put together for the answer. "If successful, the system eventually produces a response, but the response is hardly a simple recall" (Lindsay & Norman, 1972, p.380, emphasis added).

The Q/A model developed by J. R. Anderson and Bower (1974) deals with questions that require the simple retrieval of factual information from memory structures and is therefore based primarily on memory-matching routines. However, like Lindsay and Norman (1972), these authors have argued that Q/A processes are

---

8 The storage/processing tradeoff (Rumelhart & Ortony, 1977; Smith, 1977) may be a partial explanation for this feature of human information processing.

similar to problem-solving[9] strategies.

Greeno (1976) too has emphasized that, often, Q/A goes beyond comprehension. Using the example of questions requiring the explanation of phenomena and relationships, and assuming that people can have a schema for 'explanation'[10], Greeno argues that Q/A involves understanding the question, having the necessary knowledge structures, e.g., an explanation schema, using appropriate strategies to activate them, searching memory for relevant information, and performing the necessary selection and judgments to generate appropriate answers.

Norman (1972) provides data consistent with the notion that retrieval itself is a constructive process (see also diSibio, 1982; Spiro, 1977). To the question 'Where is the Empire State Building?' various answers can be given, e.g., "In the States", "In New York City", "On 34th Street", etc. The answer varies depending on one's knowledge of why the question was asked and by whom, and knowledge of the interrogator's knowledge. Otherwise, the answer would be at the wrong level and egocentric.

Norman (1972) emphasizes the importance of building a model of the questioner's world view and thus, raises some of the socio-psychological principles that operate in Q/A. Lehnert (1978, 1981) too includes this variable in her model. It must be mentioned however that question answering in instructional as

---

9 Both works have a means-ends-analysis conception of problem solving.

10 'Explanatory frames' have been postulated and described by Frederiksen (in prep.; Frederiksen & Frederiksen, 1981).

well as research data collection settings is subject to relatively little contextual variability. Often, the questions that teachers and educational researchers ask students are not genuine inquiry questions (Dillon, 1982; Lehnert, 1981). They tend to function as commands and, strictly speaking, there is no need for them to be interrogative. More often than not questions are asked for the purpose of gathering information about the student rather than about the queried subject. The interrogator is often the same person, i.e., the teacher or the experimenter, as far as the subject matter is concerned, s/he knows the answer, and the answerer is aware of this. For all these reasons, the pragmatics and the social dynamics of question answering interactions are more constrained in classroom settings than they normally are. Assessing the knowledge state of the questioner, therefore, ceases to be an important issue.

These constraints facilitate the task of the researcher and it may even be the case that building a theory of Q/A without contextual restrictions is not feasible. As Anderson and Bower (1974) have suggested, human question answering is so complex and open-ended that the attempt to build a general, all-purpose model of Q/A may be futile.

# CHAPTER III

# RATIONALE

As reviewed in the preceding chapter, adjunct-questions studies tell us that initial question answering can improve performance on subsequent question answering, research on comprehension informs us of, and analyzes the variety of structural constraints that operate in the process of discourse comprehension, and cognitive studies of the question-answering (Q/A) process reveal the complexities involved in it. The issue of the extent to which comprehension contributes to Q/A remains unresolved however. Schank and Lebowitz (1980) consider their computer program to have 'understood' a story when it can paraphrase it and/or answer questions about it. The two tasks are viewed as interchangeable ways of demonstrating comprehension. What is the relationship between paraphrasing and Q/A tasks however?

Frederiksen (in prep., C. H. Frederiksen, J. D. Frederiksen & Bracewell, in press; Bracewell, C. H. Frederiksen & J. D. Frederiksen, 1982) has developed a framework for task analysis central to which is the notion of <u>constraint</u>. Instead of dichotomizing tasks as, say, comprehension versus production, the

specific requirements of each task are analyzed to characterize it. In this framework, questions and other retrieval cues are viewed as instances of 'constraints on output' that can apply in 'discourse processing tasks'. From such a perspective therefore, the Q/A production mode is, ceteris paribus, more constrained than a free recall task.

As indicated in the preceding chapter, the process of comprehension itself can be viewed as one of structure building while satisfying multiple constraints (Collins, Brown & Larkin, 1980; Frederiksen, 1981). Ordinarily, discourse comprehension is mainly constrained by linguistic features, propositional structures, and underlying global conceptual structures, i.e., frames, as well as the knowledge structures of the comprehender and the comprehension context (Frederiksen, 1981). Question answering, on the other hand, involves at least the additional requirements of the question itself. A question explicitly sets up specific processing demands. While many of the inferences involved in comprehension are only 'invited' or 'signalled' by the text and are, therefore, optional, questions require inferences that one may not draw spontaneously. The answerer must not only understand the content of the question and its relationship to the text, but must also infer its cognitive and pragmatic[11] demands.

In a sense therefore, Q/A is more complex than comprehension because it requires the satisfaction of more constraints. From a different perspective however, it is less complex because the

---

11 As mentioned earlier however, in classroom settings, the pragmatics of Q/A is highly predictable.

interrogator has already done some of the structuring that the comprehender must do.

Free recall tasks are, in comparison, ambiguously structured and ill defined. In this sense, they are similar to Rorschach tests[12] and therefore, reflect the spontaneous processes of the language user. Recall tasks certainly do involve requirements which are not present in comprehension itself[13]. They involve, like Q/A, constraints related to the internal coherence of the production. They do affect memory structures and processes and do not include unretrievable information. However, compared to Q/A tasks, they bias the comprehender to a lesser extent. For this reason, this research is based on the assumption that recall protocols reflect largely comprehension.

The main objective of the study was to investigate the relationship between question answering and comprehension as evidenced by free recall. Subjects were therefore asked to read texts, recall them in their own words, and answer questions about them[14]. The expectation was that, while comprehension and Q/A are neither identical nor totally independent processes, the relationship between them is not strong enough for data obtained

---

12 I am indebted to Dr. Frederiksen for the analogy between free recall tasks and projective tests as well as many of the other ideas discussed here.

13 Here is an anecdotical piece of evidence: In a study other than this one where children were asked to retell stories orally 'in their own words', one grade-two child paused at the middle of the recall. When the experimenter asked him whether or not he remembered more, the child said: "I member something else but, but you said say it in your own way don't say the story dexack [sic]"!

14 Since free recall was used as an unbiasing kind of task, the procedures were presented in a fixed sequence. Order of the materials was counterbalanced however (see the next chapter).

under free recall constraints to account for responses to questions. Analogous to this hypothesis is the result reported by Kubes (1982) on the processes of knowledge integration (i.e., transfer of information across passages). She found that while comprehension helps one to integrate knowledge, it is not a sufficient condition for it.

A final task in which students had access to the reading material when responding to questions was also included in the study. The face validity and the importance of such a task becomes clear in view of the fact that assessment of comprehension is frequently done in this way. As Johnston (1981, 1982) has indicated however, this task differs considerably from Q/A in the absence of text in terms of its cognitive demands. With the constraints related to the surface structure of the text being relaxed, the task involves less demands concerning long-term memory storage and knowledge representation. For instance, rather than relying on retrieval and structuring processes, students may primarily depend on search strategies that locate the queried information within the text. Thus, in this case, the task would essentially be a looking-back task that requires searching information and operating upon it to generate an answer. Results were therefore expected to show little relationship between performance on this task and earlier ones.

The present study also included discourse frame type (Frederiksen, 1982, in prep.; C. H. Frederiksen & J. D. Frederiksen, 1981) and question type as variables. As well it examined the extent to which subject-produced propositions were

inferential or factual. If the frame and question type experimental variables are found to affect inference from and recall of text propositions in the same way in recall and Q/A tasks, this would reinforce the conclusion that similar processes are involved in both. However, if experimental variables known to influence text comprehension have different effects on responses to questions (either with the text present or without), then this would be consistent with the conclusion that different processes are involved.

Thus, we have two means of comparing comprehension (recall) and the Q/A tasks: (a) analysis of the degree to which comprehension performance predicts performance on Q/A tasks, and (b) comparison of patterns of effects of the experimental variables across tasks.

Globally, and on the basis of the current state of research and theory in the area of discourse processing, it was expected that subjects' (grade-six students) performance would be higher for 'text-based' questions than for 'frame-based' items (see Chapter 4 for more details on the two types of questions) and that, in the first two tasks, i.e., when students are not allowed to refer back to the text, responses would be more inferential than factual. Moreover, on the basis of previous research on frame types (Frederiksen, 1982, in prep.), it was expected that these children would find the narrative passage used in the study easier than the descriptive one. The inferential nature of language processing as well as the relative difficulty of descriptive discourse and frame-related questions were also

expected to appear in the patterns of interactions.

# CHAPTER IV

# METHOD

## SUBJECTS

Forty-two grade-six students participated in the study, the data obtained from twenty of whom were eventually included in the analyses. These children were students of two coeducational, private, English-speaking, elementary schools[15] in middle- and high-income residential area of Montréal.

## MATERIALS

Two school-type expository science texts (Tables 1 and 2), written for a late elementary grades audience were used[16]. The passages were written by the same author (Kohn, 1962) and, since they deal with the same topic, namely, molds, discourse content was to a large extent controlled. The two texts are quite similar in terms of length (500-600 words), and syntactic and propositional complexity. Moreover, comprehending them does not require specific prior knowledge and the comprehension of one text does not depend on the other.

One passage, The Discovery of Penicillin, is predominantly

---

15 I wish to thank Christine Laphkas and Dr. Eigil Pedersen for providing me access to schools. I am grateful to them and to Lilli Kormos for helping me in the data collection.

16 These passages were selected from Dr. Frederiksen's laboratory and are being used in ongoing research.

TABLE 1
Passage Used in the Study

THE DISCOVERY OF PENICILLIN

Alexander Fleming was at work in his laboratory at Saint Mary's Hospital in London, England. It was a warm day and the windows were open. A dish of disease germs, needed for an experiment, had been left uncovered. It looked like cloudy soup. As Fleming walked by he glanced at the dish. Something caught his eye. He looked again. There was a patch of blue-green Penicillium mold growing on the dish. BUT -instead of cloudy soup, thick with germs, there was a clear circle all around the mold. All of the germs near the mold were dead!

Fleming knew that this was something important. He scraped off the bit of mold and put it in a dish of its own with some food. The mold plant spread and the blue-green patch grew larger. When it was big enough, Fleming set to work to find out why the mold had killed the germs. After many weeks of difficult experiments, he finally managed to squeeze from the mold a few drops of brownish fluid. This remarkable fluid was the germ-killer. Fleming named it penicillin (say: pen-i-SILL-in).

Penicillin turned out to be a far greater germ-killer than anything ever known before. But it took so long to make even a drop of it! Even though it was surely a valuable drug, Fleming decided that making penicillin was not practical. He went on with his other work.

Ten years later, in 1938, Dr. Flory and Dr. Chain, of Oxford University, read about Fleming's discovery. They were looking for a new medicine that would help people who were wounded in World War II. Penicillin seemed like the perfect answer. But when they tried to make it, they found the work very slow, just as Fleming had.

They had finally managed to make a few drops of penicillin when they heard about a London policeman who was dying in a hospital. He had a blood infection for which there was no cure. Flory and Chain began to give him penicillin. The policeman improved very quickly. He was almost well when the supply of penicillin ran out. It took so long to make more that the poor policeman got sick again and died.

But by now the scientists were sure they had a real miracle drug. It seemed to cure many diseases -among them, pneumonia (say: new-MO-nia), scarlet fever and rheumatic (say: roo-MAT-ik) fever. But England was at war. The whole country was busy fighting German bombs. There was no time for penicillin in England.

Finally, Dr. Flory and his fellow-workers decided to go to the United States which was not yet at war. They came in 1941. They promptly started a search for a better Penicillium mold.

They sent a helper out every morning to buy all the moldy fruit she could find in the market. People laughed at her. They called her "Moldy Mary". One morning "Moldy Mary" went to a fruit

store. She came back with a very moldy cantaloupe. The mold on that cantaloupe was a different kind of Penicillium. It gave two hundred times as much penicillin as Fleming's original mold!

The drug companies soon found ways to make penicillin in large tanks. By 1946 there was enough to treat seven million patients a year. The age of miracle molds had come -but it had only just begun.

TABLE 2
Passage Used in the Study

MAGIC SEEDS

If you decided to start a flower garden, you would probably go to the store and buy some seeds. Then you would plant them in the ground. But if you want to start a mold garden you don't have to do nearly so much.

As a matter of fact, you don't have to do anything except get a good growing-place ready. The seeds will come by magic and will even plant themselves!

Did you say that magic is only for fairy stories? You are quite right -so let's take a close look at some molds under the microscope and see where these seeds really come from.

The first thing we notice is a big batch of threads that grow in all directions. They look a little bit like a tangled spider web.

If we can pick out one single thread and examine it carefully we see that it has branches like a Christmas tree. Each branch has a whole series of new branches. At the ends of some of these we see little round balls. They are really hollow cases and each one is chock-full of very tiny seeds called spores. The spores are the mold plant's seeds.

The threads and their cases are all colorless, but as we look we get a surprise. One of the spore cases has become ripe and it suddenly bursts open like a silent firecracker. We see hundreds of beautifully colored spores.

Now we have learned something: mold plants have no color at all. It is only their spores which make them look black or blue or pink or almost any other color.

A single spore is so small that perhaps a thousand of them would fit on the head of a pin. It is so light that it is lighter than the lightest thing you can think of. It is much, much lighter than milkweed fluff. It is lighter than the grains of dust you see in a beam of sunshine. It is so light that it can just hang in the air or float about for days and not fall down.

The tiniest breath of air, even the breeze you make when you walk through a room will make the spores dance. The storm you make when you sneeze or turn on the electric fan will send them scurrying far and wide.

Each mold growth may produce millions and millions of spores and they all float around looking for a place to grow. They are in the air all around you right now, and on the floor and on your shoes and even on your hair!

Unlike green plants, molds do not need any light but they do need air, food, and water. The mildew could never have grown on your towel if you had hung it up to dry before putting it in the hamper.

Now do you know why you don't have to plant mold seeds? Wherever there is food, air, and moisture, some mold spores will almost certainly settle and begin to grow.

narrative with respect to its frame type (Frederiksen, 1982, in prep.; C. H. Frederiksen & J. D. Frederiksen, 1981), and expresses the event sequence involved in the process of the discovery and production of Penicillin (see Table 1). According to frame theory's analysis (op. cit.), this text is moreover underlied by an implicitly encoded problem frame[17]. The other passage, Magic Seeds, is primarily of descriptive frame type (op. cit.) and deals with the structure of mold plants when viewed through a microscope and the process of mold spores' propagation (see Table 2).

The propositional structure of the two texts was analyzed[18] according to an updated and simplified version of Frederiksen's (1975) semantic representation system (Frederiksen, 1981). Tables 3 and 4 present the two analyses. In the propositional notation, head elements or predicates represent objects, actions, or interpropositional relations. Arguments, on the other hand, represent relations followed by slots filled by concepts or other propositions, or represent the propositions that head element relations connect. Propositional structures are generated by an explicit propositional grammar represented alternatively as Augmented Transition Networks or in Backus-Naur Form notation (Frederiksen, in prep.).

The experimental materials also included questions relevant to each passage. The questions were intended to be

---

17 Interestingly, this analysis is compatible with the view of those philosophers of science who conceive of scientific activity (e.g., making Penicillin) as a problem-solving process (e.g., Kuhn, 1970).
18 The analyses had been done by Dr. Frederiksen.

TABLE 3

Propositional Analysis of The Discovery of Penicillin

| NO. | PREDICATE | ARGUMENTS |
|---|---|---|
| 1.0 | WORK(BE) | PAT:ALEXANDER FLEMING=LOC:(IN)LABORATORY, TNS:PAST,ASPCT:CONT; |
| 1.1 | (POSS) | PAT:HIS,OBJ:LABORATORY; |
| 1.2 | LABORATORY | LOC:(AT)ST.MARY'S HOSPITAL; |
| 1.3 | ST.MARY'S HOSP. | LOC:(IN)LONDON ENGLAND; |
| 2.0 | DAY | ATT:WARM,TNS:PAST; |
| 3.0 | WINDOWS | ATT:OPEN,TNS:PAST; |
| 4.0 | LEAVE | OBJ:DISH,RSLT:4.1=TNS:PAST,ASPCT:COMP; |
| 4.1 | DISH | ATT:UNCOVERED; |
| 4.2 | GERMS | LOC:(OF)DISH; |
| 4.3 | GERMS | ATT:DISEASE; |
| 4.4 | NEED | OBJ:GERMS,GOAL:(FOR)EXPERIMENT(TOK); |
| 5.0 | LOOK(LIKE) | OBJ:IT,THEME:5.1=TNS:PAST; |
| 5.1 | PROX | [IT],[SOUP]; |
| 5.2 | SOUP | ATT:CLOUDY; |
| 6.0 | WALK BY | PAT:FLEMING=TNS:PAST,TEM:(AS); |
| 7.0 | GLANCE AT | PAT:HE,OBJ:DISH=TNS:PAST,TEM: ; |
| 7.1 | EQUIV(TEM): | [6.0],[7.0]; |
| 8.0 | CATCH ONE'S EYE | PAT:HIS,OBJ:SOMETHING=TNS:PAST; |
| 9.0 | LOOK | PAT:HE=TNS:PAST,ASPCT:ITER(AGAIN); |
| 10.0 | GROW | PAT:PATCH=LOC:(ON)DISH,TNS:PAST,ASPCT:CONT; |
| 10.1 | MOLD | CAT:PATCH; |
| 10.2 | MOLD | ATT:BLUE-GREEN; |
| 10.3 | MOLD | CAT:PENNICILLIUM; |
| 11.0 | SOUP | ATT:CLOUDY; |
| 11.1 | SOUP | ATT:THICK; |
| 11.2 | CAU: | [GERMS],[11.1];(WITH) |
| 11.3 | CIRCLE | LOC:(ALL AROUND)MOLD,TNS:PAST; |
| 11.4 | CIRCLE | ATT:CLEAR,TNS:PAST; |
| 12.0 | GERMS(NUM:ALL) | LOC:NEAR; |
| 12.1 | MOLD | LOC: ; |
| 12.2 | PROX(LOC): | [12.0],[12.1]; |
| 12.3 | GERMS | ATT:DEAD,TNS:PAST; |
| 13.0 | KNOW | PAT:FLEMING,THM:13.1,13.2=TNS:PAST; |
| 13.1 | IDENT: | [THIS],[SOMETHING]; |
| 13.2 | SOMETHING | ATT:IMPORTANT; |
| 14.0 | SCRAPE OFF | AGT:HE,OBJ:BIT=TNS:PAST; |
| 14.1 | MOLD | PRT:BIT; |
| 15.0 | PUT | AGT:AND,OBJ:IT,RSLT:15.1,15.2,15.3,15.4,15.5 =TNS:PAST; |
| 15.1 | IT | LOC:(IN)DISH; |
| 15.2 | FOOD | DEG:SOME; |
| 15.3 | FOOD | LOC: ; |
| 15.4 | PROX:(WITH) | [15.1],[15.3];(WITH) |
| 15.5 | (POSS)(OWN) | PAT:ITS,OBJ:DISH; |
| 16.0 | SPREAD | PAT:MOLD=TNS:PAST; |
| 16.1 | PLANT | CAT:MOLD; |
| 17.0 | GROW | AGT:PATCH,RSLT:17.1=TNS:PAST; |

TABLE 3 (cont'd)

| | | |
|---|---|---|
| 17.1 | ORD: | [17.2],[ ]; |
| 17.2 | PATCH | ATT:LARGER,DEG: ; |
| 17.3 | PATCH | ATT:BLUE-GREEN; |
| 18.0 | IT | ATT:BIG,DEG:(ENOUGH),TNS:PAST,TEM:WHEN; |
| 19.0 | SET TO WORK | PAT:FLEMING,GOAL:19.1=TNS:PAST,TEM: ; |
| 19.1 | FIND OUT | THM:19.2; |
| 19.2 | CAU: | [WHY],[19.3]; |
| 19.3 | KILL | AGT:MOLD,OBJ:GERMS=TNS:PAST,ASPCT:COMP; |
| 19.4 | EQUIV:(TEM) | [18.0],[19.0]; |
| 20.0 | EXPERIMENTS | DUR:MANY WEEKS,ATT:DIFFICULT,TEM:AFTER; |
| 20.1 | ORD:(TEM) | [20.0],[20.2]; |
| 20.2 | MANAGE | PAT:HE,ACT:20.3=TNS:PAST,ATT:FINALLY,TEM: ; |
| 20.3 | SQUEEZE | SOURCE:MOLD,RSLT:DROPS; |
| 20.4 | DROPS | NUM:FEW; |
| 20.5 | FLUID | ATT:BROWNISH; |
| 20.6 | FLUID | PRT:DROPS; |
| 21.0 | IDENT: | [FLUID],[GERM-KILLER]; |
| 21.1 | FLUID | ATT:REMARKABLE; |
| 22.0 | NAME | AGT:FLEMING,OBJ:IT,RSLT:22.1=TNS:PAST; |
| 22.1 | EQUIV: | [IT],[PENICILLIN]; |
| 23.0 | TURN OUT | OBJ:PENICILLIN,RSLT:23.1,23.2,23.3,23.4,23.5 =TNS:PAST; |
| 23.1 | GERM-KILLER | CAT:PENICILLIN; |
| 23.2 | GERM-KILLER | ATT:GREATER,DEG:FAR,DEG: ; |
| 23.3 | ORD:(DEG) | [23.2],[23,4]; |
| 23.4 | ANYTHING | ATT:(GREATER),DEG: ; |
| 23.5 | KNOW | OBJ:ANYTHING=TNS:PAST,TEM:(EVER)BEFORE; |
| 24.0 | TAKE | ACT:24.1=DUR:SO LONG,TNS:PAST; |
| 24.1 | MAKE | RSLT:DROP; |
| 24.2 | IT | PRT:DROP; |
| 25.0 | DRUG | CAT:IT,TNS:PAST; |
| 25.1 | DRUG | ATT:VALUABLE,QUAL:SURELY; |
| 26.0 | DECIDE | AGT:FLEMING,THM:26.1=TNS:PAST; |
| 26.1 | MAKE | RSLT:PENICILLIN=ATT/NEG:PRACTICAL,TNS:PAST; |
| 27.0 | GO ON WITH | PAT:HE,ACT:27.1=TNS:PAST; |
| 27.1 | OTHER | PAT:HIS; |
| 27.2 | WORK | CAT:OTHER; |
| 28.0 | READ | PAT:DR.FLORY,DR.CHAIN,THM:28.1=TNS:PAST, TEM:1938,TEM:LATER; |
| 28.1 | DISCOVERY | PAT:FLEMING; |
| 28.2 | FLORY,CHAIN | LOC:OXFORD UNIVERSITY; |
| 28.3 | DIFF=10 YEARS(LATER) | [ ],[28.0]; |
| 29.0 | LOOK FOR | PAT:THEY,GOAL:MEDICINE=TEM:PAST,ASPCT:CONT; |
| 29.1 | MEDICINE | ATT:NEW; |
| 29.2 | HELP | PAT:MEDICINE,OBJ:PEOPLE; |
| 29.3 | WOUND | OBJ:PEOPLE,ACT:29.4(IN)=TNS:PAST; |
| 29.4 | WORLD WAR II | ; |
| 30.0 | SEEM | THM:30.1=TNS:PAST; |
| 30.1 | PROX: | [PENICILLIN],[ANSWER];(LIKE) |
| 30.2 | ANSWER | ATT:PERFECT; |
| 31.0 | TRY | PAT:THEY,GOAL:31.1=TNS:PAST,TEM:WHEN; |
| 31.1 | MAKE | RSLT:IT; |
| 32.0 | FIND | AGT:THEY,THM:32.1=TNS:PAST,TEM: ; |
| 32.1 | WORK | ATT:SLOW,DEG:VERY; |

TABLE 3 (cont'd)

| | | |
|---|---|---|
| 32.2 | EQUIV:(TEM) | [31.0],[32.0]; |
| 33.0 | FIND | AGT:FLEMING,THM:32.1=TNS:PAST,ACPCT:COMP; |
| 34.0 | MANAGE | PAT:THEY,ACT:34.1=TNS:PAST,TEM:FINALLY; |
| 34.1 | MAKE | RSLT:34.2; |
| 34.2 | DROPS(NUM:A FEW) | CAT:PENICILLIN; |
| 35.0 | HEAR ABOUT | PAT:THEY,OBJ:POLICEMAN=TNS:PAST,TEM:WHEN; |
| 35.1 | EQUIV:(TEM) | [34.0],[35.0]; |
| 35.2 | POLICEMAN | LOC:LONDON; |
| 35.3 | DIE | PAT:WHO=TNS:PAST,ASPCT:CONT, LOC:(IN)HOSPITAL; |
| 36.0 | HAVE | PAT:HE,OBJ:INFECTION=TNS:PAST; |
| 36.1 | INFECTION | ATT:BLOOD; |
| 36.2 | (CURE) | AGT:NULL(NO CURE),OBJ:(FOR)INFECTION; |
| 37.0 | GIVE | AGT:FLORY,CHAIN,REC:HIM,OBJ:PENICILLIN =TNS:PAST,ASPCT:INCPT(BEGAN); |
| 38.0 | IMPROVE | PAT:POLICEMAN=ATT:QUICKLY,DEG:VERY,TNS:PAST; |
| 39.0 | HE | ATT:WELL,QUAL:ALMOST,TNS:PAST,TEM:; |
| 40.0 | RUN OUT | OBJ:SUPPLY=TNS:PAST,TEM:WHEN; |
| 40.1 | SUPPLY | CAT:PENICILLIN; |
| 40.2 | EQUIV:(TEM) | [39.0],[40.0]; |
| 41.0 | TAKE | ACT:41.1=DUR:SO LONG,TNS:PAST; |
| 41.1 | MAKE | RSLT:MORE; |
| 41.2 | GET | AGT:POLICEMAN,RSLT:41.4=TNS:PAST, ASPCT:ITER(AGAIN); |
| 41.3 | POLICEMAN | ATT:POOR; |
| 41.4 | SICK | PAT:POLICEMAN; |
| 41.5 | DIE | AGT:(AND)=TNS:PAST; |
| 41.6 | COND: | [41.0],[41.2,41.5]; |
| 42.0 | BE SURE | PAT:SCIENTISTS,THM:42.1=TNS:PAST,TEM:BY NOW; |
| 42.1 | (POSS) | PAT:THEY,OBJ:DRUG; |
| 42.2 | DRUG | CAT:MIRACLE; |
| 42.3 | DRUG | ATT:REAL; |
| 43.0 | SEEM | OBJ:IT,THM:43.1=TNS:PAST; |
| 43.1 | CURE | OBJ:DISEASES; |
| 43.2 | DISEASES(NUM:MANY) | CAT:PNEUMONIA; |
| 43.3 | DISEASES(NUM:MANY) | CAT:SCARLET FEVER; |
| 43.4 | DISEASES(NUM:MANY) | CAT:RHEUMATIC FEVER; |
| 44.0 | ENGLAND | ATT:AT WAR; |
| 45.0 | BE BUSY | PAT:COUNTRY,ACT:45.1=TNS:PAST,ASPCT:CONT; |
| 45.1 | FIGHT | OBJ:BOMBS; |
| 45.2 | BOMBS | ATT:GERMAN; |
| 45.3 | COUNTRY | ATT:WHOLE; |
| 46.0 | | OBJ:(FOR)PENICILLIN=LOC:(IN)ENGLAND; |
| 46.1 | COND: | [46.2],[46.0]; |
| 46.2 | | DUR:NULL(NO TIME); |
| 47.0 | DECIDE | AGT:DR.FLORY,FELLOW-WORKERS,GOAL:47.1 =TNS:PAST,TEM:FINALLY; |
| 47.1 | MOVE(GO) | GOAL:47.2; |
| 47.2 | | LOC:UNITED STATES; |
| 47.3 | UNITED STATES | ATT:AT WAR,NEG,TEM:YET,TNS:PAST; |
| 47.4 | (POSS) | PAT:DR.FLORY,OBJ:FELLOW-WORKERS; |
| 48.0 | MOVE(COME) | AGT:THEY=TEM:1941,TNS:PAST; |
| 49.0 | START | AGT:THEY,ACT:49.1=ATT:PROMPTLY,TNS:PAST; |
| 49.1 | SEARCH | GOAL:49.2,49.3; |

TABLE 3 (cont'd)

```
49.2  MOLD                  CAT:PENICILLIUM;
49.3  MOLD                  ATT:BETTER;
50.0  SEND OUT              PAT:THEY,OBJ:HELPER,GOAL:50.1=TEM:MORNING,
                            TNS:PAST,ASPCT:ITER(EVERY);
50.1  BUY                   OBJ:FRUIT;
50.2  FRUIT(DEF)            ATT:MOLDY;
50.3  FRUIT(DEF)            UNIV(ALL);
50.4  FIND                  AGT:SHE,OBJ:FRUIT=LOC:(IN)MARKET,
                            MOD:CAN(COULD);
51.0  LAUGH AT              PAT:PEOPLE,OBJ:HER=TNS:PAST;
52.0  CALL                  PAT:THEY,OBJ:HER,THM:'MOLDY MARY'=TNS:PAST;
53.0  GO(MOVE)              AGT:MOLDY MARY,RSLT:53.1=TEM:ONE MORNING,
                            TNS:PAST;
53.1  MOLDY MARY            LOC:STORE;
53.2  STORE                 ATT:FRUIT;
54.0  COME BACK WITH        AGT:SHE,OBJ:CANTALOUPE=TNS:PAST;
54.1  CANTALOUPE            ATT:MOLDY,DEG:VERY;
55.0  MOLD                  LOC:(ON)CANTALOUPE(THAT);
55.1  PENICILLIN            CAT:DIFFERENT KIND(TOK);
55.2  DIFFERENT KIND        CAT:55.0,TNS:PAST;
56.0  GIVE                  AGT:IT,RSLT:56.1=TNS:PAST;
56.1  PENICILLIN            ATT:MUCH,DEG:(AS);
56.2  MULT:                 [56.1],[57.3]=200;
57.0  (GIVE)                AGT:57.1,57.2,RSLT:57.3=TNS:PAST;
57.1  MOLD                  ATT:ORIGINAL;
57.2  (POSS)                PAT:FLEMING,OBJ:MOLD;
57.3  PENICILLIN            ATT:MUCH,DEG:(AS);
58.0  FIND                  AGT:COMPANIES,THM:58.1=TEM:SOON,TNS:PAST;
58.1  WAYS                  GOAL:58.2;
58.2  MAKE                  RSLT:PENICILLN=LOC:(IN)TANKS;
58.3  COMPANIES             ATT:DRUG;
58.4  TANKS                 ATT:LARGE;
59.0  TREAT                 OBJ:PATIENTS(TOK,NUM:7M)=DUR:A YEAR;
59.1  (PENICILLIN)          ATT:ENOUGH,TEM:(BY)1946,TNS:PAST;
59.2  COND:                 [59.1],[59.0];
60.0  COME                  RSLT:60.1=TNS:PAST,ASPCT:COMP;
60.1  AGE OF MIRICLE MOLDS   ;
61.0  IT                    ASPCT:INCPT,COMP,TNS:PAST;
```

TABLE 4
Propositional Analysis of Magic Seeds

| NO. | PREDICATE | ARGUMENTS |
|---|---|---|
| 1.0 | COND:(IF) | [1.1],[2.0,3.0]; |
| 1.1 | DECIDE | AGT:YOU,THM:1.2,1.3=TNS:PAST; |
| 1.2 | (START) | RSLT:1.3=ASPCT:INCPT(START); |
| 1.3 | GARDEN | ATT:FLOWER; |
| 2.0 | MOVE(GO) | AGT:YOU,RSLT:2.1=TEM: ,MOD:QUAL(PROBABLY), MOD:COND; |
| 2.1 | YOU | LOC:(TO)STORE; |
| 3.0 | BUY | PAT:(AND),OBJ:SEEDS(NUM:SOME)=TEM: ; |
| 4.0 | ORD:TEM | [2.0,3.0],[4.1]; |
| 4.1 | PLANT | AGT:YOU,OBJ:THEM,RSLT:4.2=MOD:COND; |
| 4.2 | THEM | LOC:(IN)GROUND; |
| 5.0 | COND:(IF) | [5.1],[6.0]; |
| 5.1 | WANT | PAT:YOU,GOAL:5.2,5.3=; |
| 5.2 | START | OBJ:GARDEN; |
| 5.3 | GARDEN | ATT:MOLD; |
| 6.0 | DO | PAT:YOU=ATT:MUCH,DEG:SO,MOD:ROOT(HAVE TO), NEG; |
| 6.1 | DIFF=NEARLY:DEG | [6.0],[ ]; |
| 7.0 | DO | PAT:YOU,ACT:ANYTHING=MOD:ROOT(HAVE TO),NEG; |
| 8.0 | GET READY | PAT:(EXCEPT),OBJ:PLACE=; |
| 8.1 | PLACE | ATT:GOOD; |
| 8.2 | PLACE | ATT:GROWING; |
| 9.0 | COME | OBJ:SEEDS=TNS:FUT; |
| 9.1 | MAGIC | ; |
| 9.2 | CAU:(BY) | [9.1],[9.0]; |
| 10.0 | PLANT | AGT:(AND),OBJ:THEMSELVES=TNS:FUT; |
| 11.0 | SAY | AGT:YOU,THM:11.1=TNS:PAST,INT; |
| 11.1 | PROX:(FOR) | [11.2],[11.3]; |
| 11.2 | MAGIC | ; |
| 11.3 | STORIES | ATT:FAIRY; |
| 12.0 | (BE) | PAT:YOU=ATT:RIGHT,MOD:QUAL(QUITE); |
| 13.0 | REQUEST(LET'S) | THM:13.1,13.2,14.0=; |
| 13.1 | TAKE | PAT:US,ACT:13.2=; |
| 13.2 | LOOK | OBJ:(AT)MOLDS(TOK,NUM:SOME)=ATT:CLOSE; |
| 13.3 | MOLDS | LOC:(UNDER)MICROSCOPE; |
| 14.0 | SEE | PAT:(AND),THM:14.1=; |
| 14.1 | COME | AGT:SEEDS,SOURCE:14.2=MOD:QUAL(REALLY); |
| 14.2 | SEEDS | LOC:WHERE; |
| 15.0 | NOTICE | PAT:WE,OBJ:THING=; |
| 15.1 | ORD:(FIRST) | [THING],[ ]; |
| 15.2 | IDENT: | [THING],[15.3]; |
| 15.3 | BATCH | PRT:THREADS; |
| 15.4 | BATCH | ATT:BIG; |
| 15.5 | GROW | PAT:THREADS=LOC:(DIR)ALL; |
| 16.0 | LOOK | OBJ:THEY,THM:16.1=; |
| 16.1 | PROX:ATT | [16.2],[16.3,16.4],MOD:QUAL(A LITTLE BIT); |
| 16.2 | THEY | ATT: ; |
| 16.3 | WEB(TOK) | ATT:TANGLED; |

TABLE 4 (cont'd)

| | | |
|---|---|---|
| 16.4 | WEB(TOK) | ATT:SPIDER; |
| 17.0 | COND:(IF) | [17.1,18.0],[19.0]; |
| 17.1 | PICK OUT | AGT:WE,OBJ:THREAD(NUM:ONE SINGLE)=MOD:CAN; |
| 18.0 | EXAMINE | PAT:(AND),OBJ:IT=ATT:CAREFULLY; |
| 19.0 | SEE | PAT:WE,THM:19.1,19.2=; |
| 19.1 | IT | PRT:BRANCHES; |
| 19.2 | TREE | ATT:CHRISTMAS; |
| 19.3 | PROX:(LIKE) | [BRANCHES],[19.2]; |
| 20.0 | BRANCH(TOK) | PRT:SERIES; |
| 20.1 | SERIES | ATT:WHOLE; |
| 20.2 | SERIES | PRT:BRANCHES; |
| 20.3 | BRANCHES | ATT:NEW; |
| 21.0 | SEE | PAT:WE,THM:21.1,21.2,21.3,21.4,21.5=; |
| 21.1 | THESE | TOK:SOME; |
| 21.2 | SOME | PRT:ENDS; |
| 21.3 | BALLS | LOC:(AT)ENDS; |
| 21.4 | BALLS | ATT:LITTLE; |
| 21.5 | BALLS | ATT:ROUND; |
| 22.0 | IDENT: | [THEY],[CASES],MOD:QUAL(REALLY); |
| 22.1 | CASES | ATT:HOLLOW; |
| 23.0 | ONE(TOK:EACH) | ATT:FULL,DEG:CHOCK; |
| 23.1 | SEEDS | LOC:(IN)ONE(TOK:EACH); |
| 23.2 | SEEDS | ATT:TINY,DEG:VERY; |
| 23.3 | EQUIV: | [SEEDS],[SPORES]; |
| 24.0 | IDENT: | [SPORES],[SEEDS]; |
| 24.1 | PLANT | PRT:SEEDS; |
| 24.2 | PLANT | ATT:MOLD; |
| 25.0 | ALL THREADS(GEN,UNIV) | ATT:COLORLESS; |
| 25.1 | THREADS(DEF) | GEN,UNIV:ALL THREADS; |
| 25.2 | ALL CASES(GEN,UNIV) | ATT:COLORLESS; |
| 25.3 | CASES(DEF) | GEN,UNIV:ALL CASES; |
| 25.4 | THREADS | PRT:CASES; |
| 26.0 | LOOK | PAT:WE=TEM:AS; |
| 27.0 | GET | PAT:WE,ACT:SURPRISE=TEM: ; |
| 27.1 | EQUIV:TEM | [26.0],[27.0]; |
| 28.0 | BECOME | OBJ:CASES(TOK:ONE),RSLT:28.1=ASPCT:COMP; |
| 28.1 | CASES | ATT:RIPE; |
| 28.2 | CASES | ATT:SPORE; |
| 29.0 | BURSTS OPEN | AGT:IT=ATT:SUDDENLY,TNS:PRES; |
| 29.1 | PROX:(LIKE) | [29.0],[29.2]; |
| 29.2 | FIRECRACKER | ATT:SILENT; |
| 30.0 | SEE | PAT:WE,THM:30.1,30.2=; |
| 30.1 | SPORES | NUM:HUNDREDS; |
| 30.2 | COLOR | OBJ:SPORES=ATT:BEAUTIFULLY; |
| 31.0 | LEARN | AGT:WE,THM:31.1=TEM:NOW,TNS:PAST,ASPCT:COMP; |
| 31.1 | SOMETHING | ; |
| 31.2 | IDENT: | [31.1],[32.0,32.1]; |
| 32.0 | PLANTS(GEN) | ATT:MOLD; |
| 32.1 | PLANTS(GEN) | ATT:NO COLOR,DEG:AT ALL; |
| 33.0 | MAKE | AGT:SPORES,OBJ:THEM,RSLT:33.1=; |
| 33.1 | LOOK | OBJ:THEM,THM:33.2=; |
| 33.2 | OR-ALT: | [33.3],[33.4],[33.5],[33.6]; |
| 33.3 | THEM | ATT:BLACK; |
| 33.4 | THEM | ATT:BLUE; |

TABLE 4 (cont'd)

```
33.5  THEM                    ATT:PINK;
33.6  THEM                    ATT:ANY OTHER COLOR,MOD:QUAL(ALMOST);
33.7  THEIR                   PRT:SPORES;
34.0  COND:                   [34.1],[34,2];
34.1  SPORE(TOK,NUM:SINGLE)   ATT:SMALL,DEG:SO;
34.2  FIT                     PAT:THEM(NUM:1000)=LOC:HEAD,
                              MOD:QUAL(PERHAPS);
34.3  PIN                     PRT:HEAD;
35.0  IT                      ATT:LIGHT,DEG:SO;
35.1  COND:                   [35.0],[35.2];
35.2  ORD:DEG                 [35.3],[35.4];
35.3  IT                      ATT:LIGHTER,DEG: ;
35.4  THING                   ATT:LIGHT,DEG:(ER),DEG:(EST);
35.5  ORD:SUP:DEG             [35.4],[ ];
35.6  THINK OF                PAT:YOU,OBJ:THING=MOD:CAN;
36.0  IT                      ATT:LIGHT,DEG: ;
36.1  MILKWEED FLUFF          ATT:LIGHT,DEG: ;
36.2  DIFF=2XMUCH:DEG         [36.0],[36.1];
36.3  ORD:DEG                 [36.0],[36.1];
37.0  IT                      ATT:LIGHT,DEG: ;
37.1  GRAINS                  ATT:LIGHT,DEG: ;
37.2  ORD:DEG                 [37.0],[37.1];
37.3  SEE                     PAT:YOU,OBJ:GRAINS=;
37.4  GRAINS                  LOC:(IN)BEAM;
37.5  SUNSHINE                PRT:BEAM;
37.6  DUST                    CAT:(OF)GRAINS;
38.0  COND:                   [38.1],[38.2,38.3,38.4];
38.1  IT                      ATT:LIGHT,DEG:SO;
38.2  HANG                    PAT:IT=LOC:(IN)AIR,DUR:DAYS,MOD:CAN;
38.3  FLOAT ABOUT             PAT:(OR)=DUR:DAYS;
38.4  FALL DOWN               PAT:(AND)=NEG;
39.0  CAU:                    [39.1,39.2,39.3,39.4,39.5],[39.7];
39.1  AIR                     PRT:BREATH;
39.2  BREATH                  ATT:TINY,DEG: ;
39.3  ORD:SUP:DEG             [39.2],[ ];
39.4  MAKE                    AGT:YOU,RSLT:BREEZE=TEM:WHEN;
39.5  WALK                    PAT:YOU=LOC:(THROUGH)ROOM(TOK),TEM: ;
39.6  EQUIV:TEM               [39.4],[39.5];
39.7  MAKE                    OBJ:SPORES,RSLT:39.8=;
39.8  DANCE                   PAT:SPORES=;
40.0  MAKE                    AGT:YOU,RSLT:STORM=TEM:;
40.1  SNEEZE                  PAT:YOU=TEM:WHEN;
40.2  TURN ON                 AGT:(OR),OBJ:FAN=TEM: ;
40.3  FAN                     ATT:ELECTRIC;
40.4  EQUIV:TEM               [40.0],[40.1];
40.5  EQUIV:TEM               [40.0],[40.2];
40.6  OR:ALT                  [40.1],[40.2];
40.7  SEND                    AGT:STORM,OBJ:THEM,RSLT:40.8=TNS:FUT;
40.8  SCURRY                  PAT:THEM=LOC:FAR,LOC:WIDE,ASPCT:CONT;
41.0  PRODUCE                 AGT:EACH,RSLT:41.3=MOD:QUAL(MAY);
41.1  MOLD                    PRT:GROWTH;
41.2  GROWTH                  TOK:EACH;
41.3  SPORES                  NUM:MILLIONS AND MILLIONS;
42.0  FLOAT AROUND            PAT:ALL,ACT:42.2;
```

TABLE 4 (cont'd)

| | | |
|---|---|---|
| 42.1 | THEY | GEN,UNIV:ALL; |
| 42.2 | LOOK | OBJ:PLACE,GOAL:(FOR)42.3=; |
| 42.3 | GROW | =LOC:PLACE; |
| 43.0 | THEY | LOC:(IN)AIR,TEM:RIGHT NOW; |
| 43.1 | AIR | LOC:(ALL AROUND)YOU; |
| 43.2 | THEY | LOC:(ON)FLOOR; |
| 43.3 | THEY | LOC:(ON)SHOES; |
| 43.4 | (POSS) | PAT:YOUR,OBJ:SHOES=; |
| 43.5 | THEY | LOC:(ON)HAIR; |
| 43.6 | YOUR | PRT:HAIR; |
| 44.0 | PROX:(LIKE) | [44.1],[44.3],NEG; |
| 44.1 | PLANTS | ATT:GREEN; |
| 44.2 | NEED | PAT:MOLDS,OBJ:LIGHT=NEG; |
| 44.3 | LIGHT | ATT:(AMOUNT),DEG:ANY; |
| 45.0 | NEED | PAT:THEY,OBJ:AIR,FOOD,WATER=; |
| 46.0 | COND:(IF) | [47.0,48.0,48.1],[46.1,46.2]; |
| 46.1 | GROW | PAT:MILDEW=LOC:(ON)TOWEL,TNS:PAST, ASPCT:COMP,MOD:COND,MOD:CAN,NEG(NEVER); |
| 46.2 | (POSS) | PAT:YOUR,OBJ:TOWEL=; |
| 47.0 | HANG | AGT:YOU,OBJ:IT,RSLT:47.1,GOAL:DRY =TEM:BEFORE,TNS:PAS,ASPECT:COMP; |
| 47.1 | IT | LOC:UP; |
| 48.0 | PUT | OBJ:IT,RSLT:48.2=TEM: ; |
| 48.1 | IT | LOC:(IN)HAMPER; |
| 48.2 | ORD:TEM | [47.0],[48.0]; |
| 49.0 | KNOW | PAT:YOU,THM:49.1=TEM:NOW,TNS:PRES,INT; |
| 49.1 | COND: | [?(WHY)],[49.2,49.3]; |
| 49.2 | PLANT | PAT:YOU,OBJ:SEEDS=MOD:ROOT,NEG; |
| 49.3 | SEEDS | ATT:MOLD; |
| 50.0 | PROX:LOC | [50.1],[51.0,51.1,52.0]; |
| 50.1 | AND: | [50.2],[50.3],[50.4]; |
| 50.2 | FOOD | LOC:WHEREVER; |
| 50.3 | AIR | LOC:WHEREVER; |
| 50.4 | MOISTURE | LOC:WHEREVER; |
| 51.0 | SETTLE | PAT:SPORES(TOK:SOME)=LOC:WHEREVER;TNS:FUT, MOD:QUAL(ALMOST CERTAINLY); |
| 51.1 | SPORES | ATT:MOLD; |
| 52.0 | GROW | PAT:(AND)=ASPCT:INCPT(BEGIN); |

comprehension[19] questions. Hence, their central requirements were not 'low-level' processes such as verbatim recall or 'high-level' operations such as integration to specific prior knowledge. Rather, the questions aimed primarily at determining whether the children had constructed a coherent model for the text (Johnston, 1982), both at the propositional level and at the frame level.

Two types of questions were, therefore, distinguished: 'Text-based' and 'frame-based' questions. Text-based questions operate on, and are concretely linked to the propositional base. They interrogate information that is explicitly encoded in the text base and/or require local inferences such as inferences connecting two adjacent propositions. Frame-related questions, on the other hand, operate on, and require accessing information about the non-linear structures that are instantiated by the text. They are assumed to require subjects to organize textual information at a global level. Thus, the two types of questions presumably vary in terms of the demands that they impose on subjects' cognitive processing: While text-based questions aim at eliciting primarily bottom-up processes, framing questions are intended to require processes that are more top-down.

Admittedly, the distinction is not sufficiently specified and in fact, low inter-rater reliability was obtained (mean=68.8%) in the classification of already generated questions. However, those raters who were familiar with discourse processing notions reached a much higher level of agreement

---

19 The work of Johnston (1982) which was referred to in the context of script theory is relevant here.

(mean=81.2%). Furthermore, the questions that were eventually used in the experiment were among those that reached perfect consensus by the latter group of raters with respect to the category to which they belonged.

As implied earlier, more questions than were eventually used were generated. An attempt was made to produce 'all' 'reasonable' comprehension questions. Tables 5 and 6 present the questions that were eventually presented to the students. Several criteria were used to select these questions. Face validity was the first selection criterion. A second criterion was to avoid over- and under-representation of a part of text. Inter-rater agreement with respect to classification of items was a further consideration. Finally, those questions that appeared incomprehensible or ambiguous to the students who participated in a pilot study were eliminated.

The frame-based questions queried information relevant to the most dominant frame type of each passage. Thus, for Magic Seeds, frame-related items required the use or the construction of descriptive frames (C. H. Frederiksen, 1982, in prep.; C. H. Frederiksen & J. D. Frederiksen, 1981) for the parts and attributes of mold plants and the mold spores' propagation process. Frame types that were relatively peripheral, e.g., the procedural frame instantiated by the first two paragraphs of the passage, were not interrogated. Frame-based questions relevant to The Discovery of Penicillin, on the other hand, queried information about the problems involved in the production of Penicillin and the plan structures implied in the passage.

TABLE 5
Questions relevant to The Discovery of Penicillin

Text-based questions:

1. Where was Fleming working?
2. In the dish that was in Fleming's laboratory, which germs were dead?
3. What did Dr. Flory's helper buy one morning?

Frame-based questions:

1. When Fleming saw that the germs in the dish were dead, he set to work and did many weeks of difficult experiments. Why?
2. Why did Dr. Flory and his fellow-workers come to the United States?
3. Why did Dr. Flory send a helper to buy moldy fruits?

TABLE 6
Questions relevant to Magic Seeds

Text-based questions:

1. What can you see at the ends of some of the branches?
2. How small is a spore?
3. Where can mold spores settle and grow?

Frame-based questions:

1. The passage says that the mold seeds seem to come by magic-and plant themselves. But it is not really by magic. Where do the seeds really come from?
2. When you look at molds under the microscope, what do you see? Describe the parts that you would see.
3. Molds may be black or blue or pink or almost any other color. Yet the passage says that they have no color at all. Explain why molds appear to be colored.

Presumably therefore, these questions required processing the text according to a problem frame (op. cit.). Although this frame type was not very explicit in the passage, it was thought that asking questions related to problem frame from grade-six students would be more interesting than asking them narrative frame-based questions. In fact, earlier research (Frederiksen, 1982, in prep.) has shown that elementary school children deal with narrative structures quite ably but have difficulty with problem frames and especially descriptive structures. Moreover, it was thought that recognizing the underlying problems and plans was necessary for understanding the passage in the fullest sense.

## EXPERIMENTAL PROCEDURE

The experiment was conducted during regular class hours. Each of the two classes which participated in the study was tested in two fifty-minute sessions, one session for each text. The optimal duration of the experiment was determined through the pilot study. In the case of each class, the two sessions were a week apart. The decision to test entire classes rather than individual students or small groups was not merely due to material considerations. The experience of the pilot study which was conducted in groups of two to four had suggested that being tested together in their own class would be a more natural situation for the children. For similar reasons, it was decided to allow students to pace their work themselves; the only time constraint was to finish all tasks at the end of the session.

All the experimental materials were provided to the students in written form and they were asked to write their answers. Since

written composition is not an easy task for many students -and in fact, children tended to produce short recalls and responses, written testing presents, from a cognitive psychology point of view, obvious disadvantages over oral testing. From an educational research perspective however, written tests do have an advantage: Most often, classroom tests are written; therefore, while students are not used to be tested orally, they find written tests quite natural. Students were told however not to "worry about things like spelling and punctuation". An approximate script of the instructions provided to children is presented in the Appendix.

The task involved reading a passage, giving a written recall for it, answering a series of questions relevant to it without access to the passage, and answering the same questions again while using the text. The sequence of the tasks was not varied because the rationale of the study required a given order. Children received two booklets at the beginning of each session. The following instructions were written on the booklets:

Booklet 1:
-Please read this passage carefully to yourself once.
-What do you remember about the passage? Please write it down in your own words Don't worry about spelling.Just write down what you remember. Do not look back at the passage.
-Please write your answer in the space below each question. Answer the questions one after the other. Do not look back at the passage.
Booklet 2:
-Here are the same questions. Please answer them again one after theother. But now you can look back at the passage and use it whileyou are answering the questions. But please don't look at your answers in the first booklet.

For methodological reasons, order of presentation of the two types of questions was counterbalanced[20]. In fact, it is reasonable to assume that answers can vary as the sequence of questions changes. Children were therefore randomly assigned to two groups of question type sequence. In one group, frame-based questions were presented first; in the second group, text-based questions appeared first. Within each question type, the order of questions followed that of the propositional base.

This within-and-between-subjects, repeated-measures design may be diagrammed as in Figure 1.

## CODING PROCEDURES

Prior to the coding of the protocols, twenty of them were selected[21]. Data obtained from three students were excluded because they were absent in one of the sessions. Nine children considered English their second language; their responses were excluded too. Of the remaining thirty protocols, four were exceptionally short, i.e., less than forty words, and were not coded and analyzed[22]. It was thought that the students who produced short recalls were not sufficiently engaged in the task for their responses to be considered representative of their

---

20 Passage order was also counterbalanced. However this variable was not included in the design as a factor. Instead, the two texts were used on two separate days. Introducing a 'rest' between the two sets of tasks was thought to decrease possible carry-over effects.

21 An N of twenty (ten in each group) was chosen partly due to material and statistical reasons and partly arbitrarily.

22 The admittedly crude outlier criterion of forty words is not as arbitrary as it may seem; protocols that were longer were at least twice as much longer.

| Between-Ss Factor / Within-Ss Factors | DISCOVERY OF PENICILLIN | | | | | | | | | | MAGIC SEEDS | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | FREE RECALL | | Q/A no TEXT | | | | Q/A with TEXT | | | | FREE RECALL | | Q/A no TEXT | | | | Q/A with TEXT | | | |
| | | | FBQ | | TBQ | | FBQ | | TBQ | | | | FBQ | | TBQ | | FBQ | | TBQ | |
| | R | I | R | I | R | I | R | I | R | I | R | I | R | I | R | I | R | I | R | I |
| Question-Type Sequence 1 (n=10) | | | | | | | | | | | | | | | | | | | | |
| Question-Type Sequence 2 (n=10) | | | | | | | | | | | | | | | | | | | | |

FIGURE 1. The Experimental Design
(FBQ: Frame-based questions, TBQ: Text-based questions; R: Recall, I: Inference)

abilities. Finally, twenty protocols were randomly selected from the remaining data.

Protocols were segmented into syntactic units using procedures based on Winograd's (1972) descriptive clausal grammar (see Bracewell, C. H. Frederiksen & J. D. Frederiksen, 1982). The following clause types are distinguished in this system: Major simple clauses: Imperatives, declaratives, interrogatives; secondary simple clauses: Adjuncts (bound adjunct, to adjunct, -ing adjunct), rank-shifted qualifiers (-en qualifier, to qualifier, -ing qualifier, rank-shifted wh-phrase), rank-shifted noun groups (rank-shifted wh-phrase, -ing group, to group, report group), complex clauses: Subject fork, verb fork. In the segmentation procedure used, all major clauses as well as 'bound adjuncts' form seperate segments. The same procedures had been used to segment the original passages -hence the numbers appearing at the left of the propositional analyses (Table 3 and 4).

Each segment of a recall protocol was numbered and coded against the propositional analysis of the relevant passage. Coding was done on the basis of the scheme developed by J. D. Frederiksen and C. H. Frederiksen (C. H. Frederiksen, in prep.). The scheme is concerned with the nature of the linkage that exists between author-produced and reader-produced propositions and essentially consists of rules for deciding whether a proposition is recalled, inferred, or both. A subject-produced segment that involves a change in the meaning of the presented proposition would be marked as inference. Inference is a broad

category covering all changes that comprehenders make to the text base, irrespective of the logical or pragmatic validity of the changes. Verbatim or paraphrased segments would be coded as recall. More formally, a stative is considered to be recalled if at least a concept-relation-concept triple is produced by the subject. An event is recalled if at least the head element, one case relation, and a concept are present in the segment. Since the recall rules allow the use of synonyms, and since in the case of the head element of events, concepts that have a superordinate-subordinate relation to each other can be used, a list of synonyms and super/subordinate terms was kept for consistency purposes. This coding scheme has reached high levels of reliability (about 90%).

The same procedures were used to code the answers, thus increasing the comparability of the data. However, while the recalls were coded against the entire passages, answers were marked against the propositions that were considered 'relevant' to the questions. These were selected apriori but some adjustments were made during the coding. The relevancy criterion was thus determined both in a top-down and a bottom-up fashion.

In the influential work of Grice (1975), being relevant is regarded as one of the basic rules for cooperative communication, in the framework of Lucas and McConkie (1980), propositions "judged to pertain in any way to answering the question" (p. 135) provide the basis for analyzing questions, and in Lehnert's (1978) computer program, relevancy is one of the most important criteria for generating good answers. Admittedly however, the

relevancy criterion is not sufficient for consistently discriminating 'right' or 'good' answers from 'wrong' or 'bad' answers. During the coding, however, it was rarely, if at all, felt that good answers obtain low scores or vice versa.

# CHAPTER V

# RESULTS

In order to perform statistical analyses, numeric data were obtained by counting the frequency of text propositions that students recalled or inferred during each task. Rather than using absolute values, percentages of propositions produced by students (relative to the number of propositions relevant in each case) were used to ensure comparability of scores across passages, tasks, and question types.

Since the study employed a repeated-measures design, and since none of the variables had more than two levels, all analyses were univariate. However, to avoid unrealistic assumptions about the structure of the covariance matrix, all analyses were conducted under a multivariate sampling model, using the Multivariance VI computer program.

Three sets of analyses were conducted: First, the amount of information children recalled or inferred during the free recall task was subjected to repeated-measures analysis of variance procedures. Next, the same techniques were used to analyze the data obtained during the question answering task (without access to text). Moreover, multiple regression analysis and analysis of covariance procedures were used to determine the predictability of these data from the free recall scores. Finally, the same

procedures were applied to data obtained from students' responses to questions when they had access to the reading material.

## THE FREE RECALL TASK

Table 7 presents the means of propositions recalled or inferred during free recall and Table 8 shows these means pooled over the between-groups factor, Sequence of Question Type (frame-based first versus text-based first). The error correlation matrix of data obtained under free recall requirements is presented in Table 9. Repeated-measures ANOVA was performed to investigate the effects of the between-groups factor, namely, Sequence of Question Type, the within-subjects factors, namely, Passage (Discovery of Penicillin, Magic Seeds) and Modality of Response (recall, inference), their interaction, and the interactions of the between-groups factor with the within-subjects factors. Although the free recall task was prior to the Q/A tasks, the Sequence of Question Type, between-subjects factor was included in this analysis in order to provide an empirical check for the random assignment of children to the two sequence groups. Table 10 summarizes the results of this analysis.

As reflected in Table 10, the main effect of Sequence of Question Type and its interactions with the other variables were not significant thus suggesting that the two groups of students were most probably quite 'similar' in the first place. High statistical significance was however reached by Passage ($F(1,18)=47.983$, $p<.0001$). The means (Tables 7 and 8) indicate that free recall performance was higher for Discovery of

TABLE 7

Mean Percent of Propositions Recalled or Inferred
for two Passages
during the Free Recall Task

| SEQUENCE GROUPS [1] | MODALITY OF RESPONSE: Recall | MODALITY OF RESPONSE: Inference |
|---|---|---|
| **Discovery of Penicillin** | | |
| 1 | 11.277 | 21.612 |
| 2 | 12.167 | 19.390. |
| **Magic Seeds** | | |
| 1 | 9.415 | 11.861 |
| 2. | 7.925 | 11.755 |

1 n=10 subjects per group.
Group 1 refers to children who received frame-based questions first.
Group 2 refers to those who received text-based questions first.

TABLE 8

Mean Percent of Propositions Recalled or Inferred during the Free Recall Task, for two Passages, Pooled over Sequence Groups

| MODALITY OF RESPONSE | PASSAGE | | |
|---|---|---|---|
| | Discovery of Penicillin | Magic Seeds | Pooled over Passage |
| Recall | 11.722[c] | 8.670[c] | 10.196[b] |
| Inference | 20.501[c] | 11.808[c] | 16.155[b] |
| Pooled over Modality | 16.112[a] | 10.239[a] | 13.176 |

a Passage effect: $F(1,18)=47.983$, $p<.0001$
b Modality effect: $F(1,18)=197.751$, $p<.0001$
c Passage x Modality interaction: $F(1,18)=19.831$, $p<.0004$

TABLE 9

Error Correlation Matrix for Percentages of Propositions Recalled or Inferred during the Free Recall Task

| | P1-R | P1-I | P2-R | P2-I |
|---|---|---|---|---|
| P1-R | 1.000 | | | |
| P1-I | 0.801 | 1.000 | | |
| P2-R | 0.691 | 0.675 | 1.000 | |
| P2-I | 0.712 | 0.539 | 0.786 | 1.000 |

(P1: Passage one, Discovery of Penicillin, P2: Passage two, Magic Seeds; R: Recall, I: Inference)

TABLE 10

Repeated-Measures Analysis of Variance of Percentages of Propositions Recalled or Inferred during the Free Recall Task

| SOURCE | F(1,18) | p |
|---|---|---|
| Between Ss: | | |
| Sequence of Question Type (S) | 0.128 | <.7247 |
| Within Ss: | | |
| Passage (P) | 47.983 | <.0001* |
| Modality of Response (M) | 197.751 | <.0001* |
| PxM | 19.831 | <.0004* |
| Interactions of Between and Within Ss Sources: | | |
| SxP | 0.006 | <.9389 |
| SxM | 1.040 | <.3215 |
| SxPxM | 3.149 | <.0929 |

*Significant at .01 level.

Penicillin than for Magic Seeds. The effect of Modality of Response was also significant ($F(1,18)=197.751$, $p<.0001$); children inferred more than they recalled. Modality, however, also interacted significantly with Passage ($F(1,18)=19.831$, $p<.0004$). As illustrated in Figure 2, while the level of recalled propositions was fairly stable across the two passages, there was much more inference in response to Discovery of Penicillin than to Magic Seeds.

## QUESTION ANSWERING WITHOUT ACCESS TO TEXT

The means of data obtained from the Q/A-no-text task, i.e., percentages of question-relevant propositions recalled or inferred, are shown in Table 11, and Table 12 presents the means pooled over Sequence groups. The results of the repeated-measures ANOVA to which these data were subjected are summarized in Table 13.

As shown in Table 13, the effect of the control variable, i.e., Sequence of Question Type, as well as its interactions with the other variables, were not significant. Most probably therefore, the order of presentation of questions has not affected students' responses to questions in the absence of text. The effect of Modality of Response was also below significance level. Thus, the overall amount of recalled propositions and that of inferred propositions did not differ significantly in this task.

The effect of Passage however was once again highly significant ($F(1,18)=25.713$, $p<.0001$). The means (Tables 11 and 12) suggest that it was easier for the children to answer

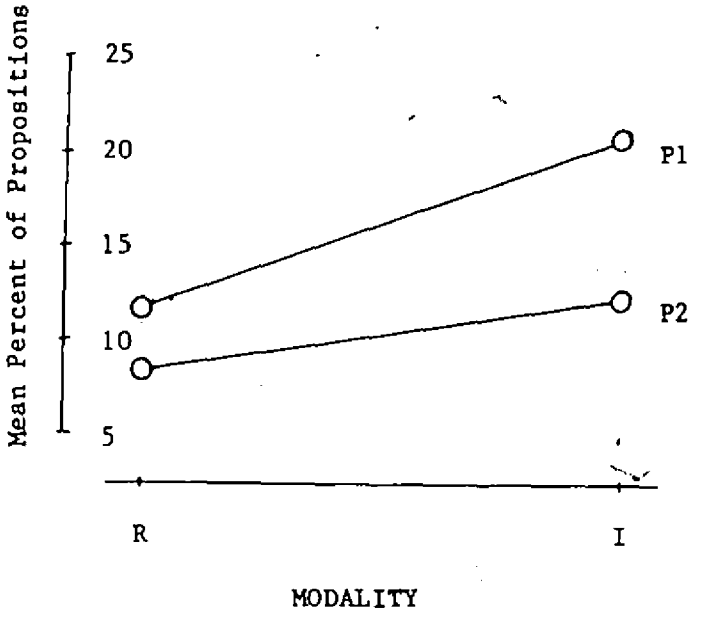


FIGURE 2: Mean Percent of Propositions Recalled or Inferred during the Free Recall Task, by Passage, Pooled over Sequence Groups (R: Recall, I: Inference; P1: Discovery of Penicillin, P2: Magic Seeds)

TABLE 11

Mean Percent of Propositions Recalled or Inferred
for two Passages and two Question Types
during the Q/A-no-Text Task

| SEQUENCE | QUESTION TYPE | | | |
|---|---|---|---|---|
| | Frame-based | | Text-based | |
| GROUPS [1] | MODALITY | | MODALITY | |
| | Recall | Inference | Recall | Inference |
| | Discovery of Penicillin | | | |
| 1 | 3.954 | 14.885 | 22.000 | 10.500 |
| 2 | 3.025 | 11.628 | 19.000 | 10.000 |
| | Magic Seeds | | | |
| 1 | 2.001 | 5.129 | 10.910 | 11.818 |
| 2 | 1.689 | 4.690 | 9.546 | 11.365 |

1 n=10 subjects per group.
Group 1 refers to children who received frame-based questions first.
Group 2 refers to those who received text-based questions first.

TABLE 12

Mean Percent of Propositions Recalled or Inferred
for two Passages and two Question Types
during the Q/A-no-Text Task,
Pooled over Sequence Groups

| MODALITY OF RESPONSE | QUESTION TYPE: Frame-based | QUESTION TYPE: Text-based | Pooled over Question Type |
|---|---|---|---|
| **Discovery of Penicillin** | | | |
| Recall | 3.490[d] | 20.500[d] | 11.995 |
| Inference | 13.257[d] | 10.250[d] | 11.754 |
| Pooled over Modality | 8.374 | 15.375 | 11.875[a] |
| **Magic Seeds** | | | |
| Recall | 1.845[d] | 10.228[d] | 6.037 |
| Inference | 4.910[d] | 11.592[d] | 8.251 |
| Pooled over Modality | 3.378 | 10.910 | 7.144[a] |
| **Pooled over Passage** | | | |
| Recall | 2.668[c] | 15.364[c] | 9.016 |
| Inference | 9.084[c] | 10.921[c] | 10.003 |
| Pooled over Modality | 5.876[b] | 13.143[b] | 9.510 |

a Passage effect: $F(1,18)=25.713$, $p<.0001$
b Question Type effect: $F(1,18)=56.533$, $p<.0001$
c Modality x Question Type interaction: $F(1,18)=42.327$, $p<.0001$
d Passage x Modality x Question Type interaction: $F(1,18)=19.951$, $p<.0003$

TABLE 13

Repeated-Measures Analysis of Variance of Percentages of Propositions Recalled or Inferred during the Q/A-no-Text Task

| SOURCE | F(1,18) | p |
|---|---|---|
| Between Ss: | | |
| Sequence of Question Type (S) | 0.519 | <.4807 |
| Within Ss: | | |
| Passage (P) | 25.713 | <.0001* |
| Modality of Response (M) | 1.544 | <.2300 |
| Question Type (Q) | 56.533 | <.0001* |
| PxM | 1.906 | <.1843 |
| PxQ | 0.082 | <.7778 |
| MxQ | 42.327 | <.0001* |
| PxMxQ | 19.951 | <.0003* |
| Interactions of Between and Within Ss Sources: | | |
| SxP | 0.470 | <.5016 |
| SxM | 0.023 | <.8820 |
| SxQ | 0.002 | <.9614 |
| SxPxM | 0.007 | <.9324 |
| SxPxQ | 0.056 | <.8157 |
| SxMxQ | 0.772 | <.3912 |
| SxPxMxQ | 0.214 | <.6496 |

*Significant at .01 level.

questions relevant to Discovery of Penicillin than Magic Seeds. The effect of Question Type was also highly significant ($F(1,18)=56.533$, $p<.0001$). Students produced more relevant answers to text-based questions than they did to frame-related items.

Statistical significance was also reached by the interaction of Question Type and Response Mode ($F(1,18)=42.327$, $p<.0001$). Figure 3 illustrates this interaction. It is found that in response to frame-based questions, children inferred much more than they recalled. In contrast, text-based questions elicited more factual ('recall') responses.

The triple interaction of Passage by Modality by Question Type was also significant ($F(1,18)=19.951$, $p<.0003$). As reflected in the tables of means (Table 11 and 12) and in Figure 4, the pattern of the Modality by Question Type interaction is not the same for both passages. For Discovery of Penicillin, the two types of questions differ markedly with respect to the mode of response that they elicit. This difference does not occur in the case of Magic Seeds.

## Analysis of the Relationship between Free Recall and Question Answering Without Text

Using the free recall data as predictors, the Q/A-no-text scores were subjected to regression analysis in order to investigate the relationship between free recall and question answering. Table 14 presents the multiple R's associated with the dependent variables, and Table 15 shows those related to the effects as derived dependent variables.

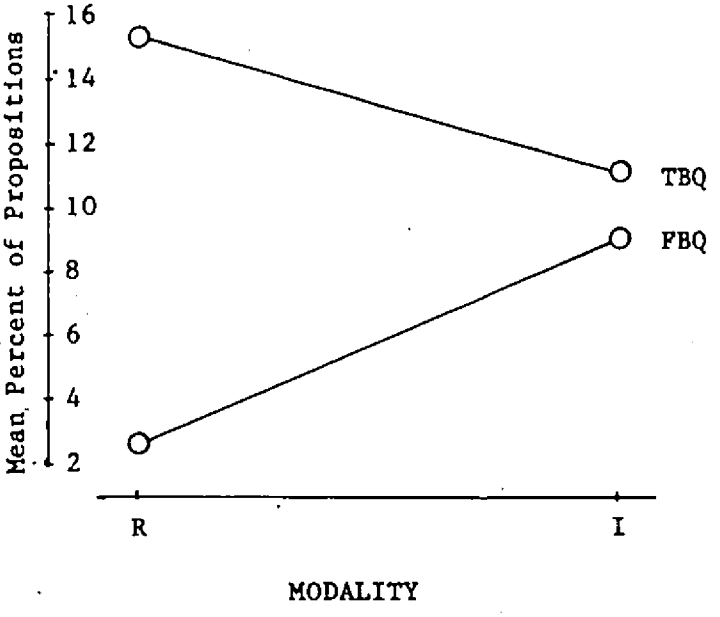


FIGURE 3: Mean Percent of Propositions Recalled or Inferred during the Q/A-no-Text Task, by Question Type, Pooled over Sequence Groups (R: Recall, I: Inference; FBQ: Frame-based Questions, TBQ: Text-based Questions)

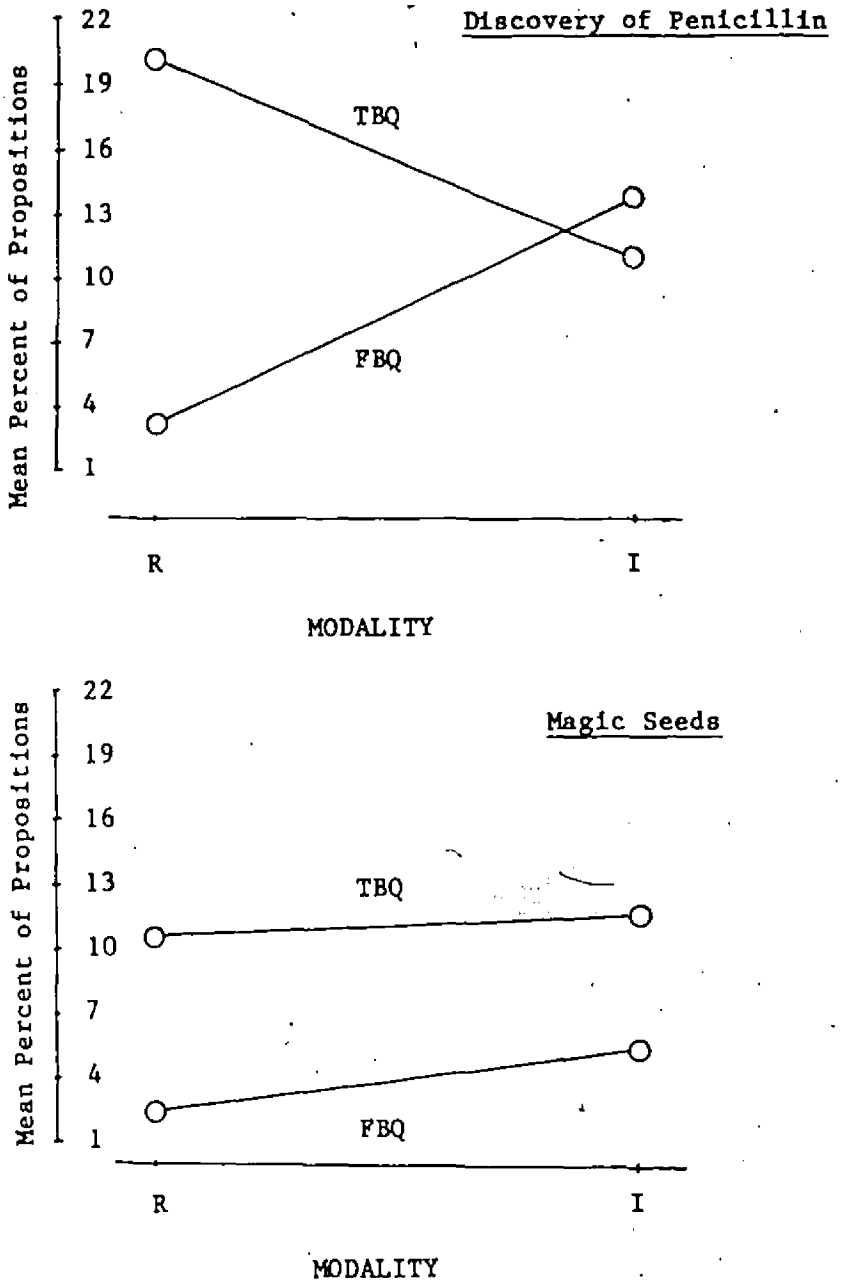


FIGURE 4: Mean Percent of Propositions Recalled or Inferred during the Q/A-no-Text Task, by Question Type and each Passage, Pooled over Sequence Groups (R: Recall, I: Inference; FBQ: Frame-based Questions, TBQ: Text-based Questions)

TABLE 14

Regression Analysis of Percentages of Propositions Recalled or Inferred during the Q/A-no-Text Task with Free Recall Task Data as Predictors 1

| DEPENDENT VARIABLES | MULTIPLE R | F(4,14) | p |
|---|---|---|---|
| P1-FBQ-R[2] | .351 | 0.493 | <.7412 |
| P1-FBQ-I | .512 | 1.241 | <.3386 |
| P1-TBQ-R[3] | .728 | 3.951 | <.0238[a] |
| P1-TBQ-I | .310 | 0.372 | <.8251 |
| P2-FBQ-R | .782 | 5.512 | <.0071[b] |
| P2-FBQ-I | .524 | 1.323 | <.3096 |
| P2-TBQ-R | .732 | 4.036 | <.0222[c] |
| P2-TBQ-I | .671 | 2.872 | <.0627[d] |

1 Four predictors: % of propositions recalled (R) or inferred (I) during the free recall task for two passages (P1: Discovery of Penicillin, P2: Magic Seeds).

2 FBQ: Frame-based questions.

3 TBQ: Text-based questions.

a Standardized weights: .501 with P1-R as predictor; -.765 with P1-I as predictor; .217 with P2-R as predictor; .476 with P2-I as predictor.

b Standardized weights: .456 with P1-R as predictor; -.735 with P1-I as predictor; .642 with P2-R as predictor; .207 with P2-I as predictor.

c Standardized weights: .779 with P1-R as predictor; -.362 with P1-I as predictor; .429 with P2-R as predictor; -.149 with P2-I as predictor.

d Standardized weights: -.335 with P1-R as predictor; -.122 with P1-I as predictor; .130 with P2-R as predictor; .805 with P2-I as predictor.

TABLE 15

Regression Analysis of the Factors involved in the Q/A-no-Text Task with Free Recall Task Data as Predictors[1]

| EFFECTS AS DEPENDENT VARIABLES | MULTIPLE R | F(4,14) | p |
|---|---|---|---|
| Total | .745 | 4.369 | <.0169[a] |
| Passage effect (P) | .438 | 0.831 | <.5272 |
| Question Type effect (Q) | .639 | 2.414 | <.0982[b] |
| Modality effect (M) | .435 | 0.817 | <.5354 |
| PxQ | .402 | 0.674 | <.6211 |
| PxM | .532 | 1.378 | <.2915 |
| QxM | .534 | 1.393 | <.2867 |
| PxQxM | .496 | 1.143 | <.3766 |

1 Four predictors: % of propositions recalled (R) or inferred (I) during the free recall task for two passages (P1: Discovery of Penicillin, P2: Magic Seeds).

a Standardized weights: .409 with P1-R as predictor; -.609 with P1-I as predictor; .236 with P2-R as predictor; .517 with P2-I as predictor.

b Standardized weights: -.693 with P1-R as predictor; .637 with P1-I as predictor; -.385 with P2-R as predictor; -.046 with P2-I as predictor.

Results suggest that total scores obtained by students in answering questions without access to text are significantly predictable from their performance in free recall ($p<.0169$). With one exception [23], this relationship is positive, i.e., the higher the performance in free recall, the more relevant are the answers. Free recall performance also accounts for three (or four, depending on the significance level that one selects) of the eight scores students obtained during the Q/A-no-text task (see Table 14). It also appears that, to some extent ($p<.0982$), free recall performance predicts the effect of Question Type when text is not accessible to students.

To investigate further the degree of similarity or difference that exists between question answering (without text) and comprehension as evidenced by free recall, repeated-measures analysis of covariance techniques were used. Results are summarized in the right-hand part of Table 16. The left-hand columns of the table repeat the results that were reported earlier in Table 13. One can thus examine the changes produced in the results by the introduction of the four covariates, namely, percent of propositions recalled or inferred during the free recall task for two passages. Table 17 presents the means prior to the introduction of the covariates and the means adjusted for the effects of the covariates.

---

23 The exception concerns the relationship between Q/A-no-text and percentage of propositions inferred for Discovery of Penicillin (P1-I) during free recall. In view of the fact that all free recall variables were correlated with each other quite strongly and positively (see Table 9), how to interpret the mentioned exception does not seem clear at this point.

TABLE 16

Repeated-Measures Analysis of Variance and Covariance of Percentages of Propositions Recalled or Inferred during the Q/A-no-Text Task

| SOURCE | WITHOUT COVARIATES F(1,18) | p | WITH COVARIATES[1] F(1,14) | p |
|---|---|---|---|---|
| Between Ss: | | | | |
| Sequence of Question Type (S) | 0.519 | <.4807 | 1.919 | <.1877 |
| Within Ss: | | | | |
| Passage (P) | 25.713 | <.0001* | 1.279 | <.2771 |
| Modality of Response (M) | 1.544 | <.2300 | 0.128 | <.7255 |
| Question Type (Q) | 56.533 | <.0001* | 2.626 | <.1275 |
| PxM | 1.906 | <.1843 | 0.028 | <.8708 |
| PxQ | 0.082 | <.7778 | 0.792 | <.3887 |
| MxQ | 42.327 | <.0001* | 1.662 | <.2183 |
| PxMxQ | 19.951 | <.0003* | 0.396 | <.5393 |
| Interactions of Between and Within Ss Sources: | | | | |
| SxP | 0.470 | <.5016 | 1.152 | <.3013 |
| SxM | 0.023 | <.8820 | 0.391 | <.5421 |
| SxQ | 0.002 | <.9614 | 0.313 | <.5850 |
| SxPxM | 0.007 | <.9324 | 0.049 | <.8279 |
| SxPxQ | 0.056 | <.8157 | 0.137 | <.7169 |
| SxMxQ | 0.772 | <.3912 | 1.365 | <.2623 |
| SxPxMxQ | 0.214 | <.6496 | 0.140 | <.7143 |

1 Four covariates: % of propositions recalled or inferred during the free recall task for two passages.

* Significant at .01 level.

TABLE 17

Unadjusted and Adjusted[1]
Mean Percent of Propositions Recalled or Inferred
for two Passages and two Question Types
during the Q/A-no-Text Task

| SEQUENCE GROUPS[2] | QUESTION TYPE: Frame-based — MODALITY: Recall — WITHOUT COVARIATES | Frame-based — Recall — WITH COVARIATES | Frame-based — Inference — WITHOUT COVARIATES | Frame-based — Inference — WITH COVARIATES | Text-based — Recall — WITHOUT COVARIATES | Text-based — Recall — WITH COVARIATES | Text-based — Inference — WITHOUT COVARIATES | Text-based — Inference — WITH COVARIATES |
|---|---|---|---|---|---|---|---|---|
| **Discovery of Penicillin** | | | | | | | | |
| 1 | 3.954 | 4.290 | 14.885 | 14.742 | 22.000 | 23.677 | 10.500 | 11.196 |
| 2 | 3.025 | 2.689 | 11.628 | 11.772 | 19.000 | 17.324 | 10.000 | 9.304 |
| **Magic Seeds** | | | | | | | | |
| 1 | 2.001 | 2.116 | 5.129 | 5.447 | 10.910 | 11.247 | 11.818 | 11.638 |
| 2 | 1.689 | 1.574 | 4.690 | 4.372 | 9.546 | 9.210 | 11.365 | 11.545 |

1 Means adjusted for the effects of four covariates: % of propositions recalled or inferred during the free recall task for two passages.

2 n=10 subjects per group.
Group 1 refers to children who received frame-based questions first.
Group 2 refers to those who received text-based questions first.

As reflected in Table 16, once the recall data are covaried out, all of the previously significant effects and interactions obtain chance level probabilities. The Passage effect which had a p-value of less than .0001 before, becomes non-significant ($p<.2771$) in the analysis of covariance. The same pattern exists in the case of the Question Type effect and the Modality by Question Type interaction. The introduction of the four covariates also greatly reduces the significance of the triple interaction of Passage by Modality by Question Type, the F-ratio of which goes below one.

It thus appears that the pattern of effects obtained from the Q/A-no-Text task can be explained to a very large extent by the free recall data.

## QUESTION ANSWERING IN THE PRESENCE OF TEXT

The means of data obtained during the Q/A-with-text task are presented in Table 18, and Table 19 shows the means pooled over Sequence groups. These data were compared to those obtained during the Q/A-no-text task. Task effect was highly significant ($F(1,18)=33.191$, $p<.0001$) with, as it might be expected, Q/A-with-text performance (mean=12.917) higher than Q/A-no-text scores (mean=9.511)[24].

Repeated-measures ANOVA procedures were applied to the Q/A-with-text data. Table 20 includes the summary of results. Once again, Sequence of Question Type had no significant effect on

---

24 Johnston (1982) too has found higher performance when students have access to the text. Opposite, and therefore counter-intuitive results however have been reported by Nicholson, Pearson, and Dykstra (1979).

TABLE 18

Mean Percent of Propositions Recalled or Inferred
for two Passages and two Question Types
during the Q/A-with-Text Task

| SEQUENCE | QUESTION TYPE | | | |
|---|---|---|---|---|
| | Frame-based | | Text-based | |
| GROUPS [1] | MODALITY | | MODALITY | |
| | Recall | Inference | Recall | Inference |
| | Discovery of Penicillin | | | |
| 1 | 11.628 | 13.023 | 34.000 | 6.500 |
| 2 | 13.721 | 8.604 | 34.500 | 4.500 |
| | Magic Seeds | | | |
| 1 | 7.065 | 5.066 | 19.998 | 8.636 |
| 2 | 7.690 | 5.127 | 20.455 | 6.136 |

1 n=10 subjects per group.
Group 1 refers to children who received frame-based questions first.
Group 2 refers to those who received text-based questions first.

TABLE 19

Mean Percent of Propositions Recalled or Inferred
for two Passages and two Question Types
during the Q/A-with-Text Task,
Pooled over Sequence Groups

| MODALITY OF RESPONSE | QUESTION TYPE: Frame-based | Text-based | Pooled over Question Type |
|---|---|---|---|
| **Discovery of Penicillin** | | | |
| Recall | 12.675[f] | 34.250[f] | 23.463[d] |
| Inference | 10.814[f] | 5.500[f] | 8.157[d] |
| Pooled over Modality | 11.745 | 19.875 | 15.810[a] |
| **Magic Seeds** | | | |
| Recall | 7.378[f] | 20.227[f] | 13.803[d] |
| Inference | 5.097[f] | 7.386[f] | 6.242[d] |
| Pooled over Modality | 6.238 | 13.807 | 10.023[a] |
| **Pooled over Passage** | | | |
| Recall | 10.027[e] | 27.239[e] | 18.633[b] |
| Inference | 7.956[e] | 6.443[e] | 7.200[b] |
| Pooled over Modality | 8.992[c] | 16.841[c] | 12.917 |

a Passage effect: $F(1,18)=32.231$, $p<.0001$
b Modality effect: $F(1,18)=78.364$, $p<.0001$
c Question Type effect: $F(1,18)=104.755$, $p<.0001$
d Passage x Modality interaction: $F(1,18)=8.485$, $p<.0093$
e Modality x Question Type interaction: $F(1,18)=72.342$, $p<.0001$
f Passage x Modality x Question Type interaction: $F(1,18)=18.505$, $p<.0005$

TABLE 20

Repeated-Measures Analysis of Variance of Percentages of Propositions Recalled or Inferred during the Q/A-with-Text Task

| SOURCE | F(1,18) | p |
|---|---|---|
| Between Ss: | | |
| Sequence of Question Type (S) | 0.236 | <.6329 |
| Within Ss: | | |
| Passage (P) | 32.231 | <.0001* |
| Modality of Response (M) | 78.364 | <.0001* |
| Question Type (Q) | 104.755 | <.0001* |
| PxM | 8.485 | <.0093* |
| PxQ | 0.086 | <.7732 |
| MxQ | 72.342 | <.0001* |
| PxMxQ | 18.505 | <.0005* |
| Interactions of Between and Within Ss Sources: | | |
| SxP | 0.092 | <.7656 |
| SxM | 1.471 | <.2409 |
| SxQ | 0.096 | <.7601 |
| SxPxM | 0.267 | <.6120 |
| SxPxQ | 0.214 | <.6490 |
| SxMxQ | 0.034 | <.8562 |
| SxPxMxQ | 0.712 | <.4100 |

*Significant at .01 level.

children's responses and it did not significantly interact with the other factors. All other main effects, however, reached significance level. Depending on Passage, students' answers in the presence of text varied significantly ($F(1,18)=32.231$, $p<.0001$). Here again, Magic Seeds was the more difficult passage (see Table 19). Modality of Response was also significant ($F(1,18)=78.364$, $p<.0001$). As it might be expected, when the original text is available to students during Q/A, their responses are more factual ('recall') than inferential. The effect of Question Type was highly significant too ($F(1,18)=104.755$, $p<.0001$). Text-based questions were easier to answer than frame-based questions.

The two-way interaction of Passage and Modality reached statistical significance ($F(1,18)=8.485$, $p<.0093$). Means relevant to this interaction are plotted in Figure 5. Answers were much more factual ('recall') for Discovery of Penicillin than for Magic Seeds. This difference however decreases in the case of inferential responses.

The Response Modality by Question Type interaction was also significant ($F(1,18)=72.342$, $p<.0001$). This interaction is illustrated in Figure 6. It appears that, while the amount of inferential answers does not vary substantially as a function of Question Type, much more factual ('recall') responses are given to text-based questions than to frame-related items.

The passage by Modality by Question Type triple interaction reached significance level ($F(1,18)=18.505$, $p<.0005$). The means relevant to this interaction are plotted in Figure 7 which

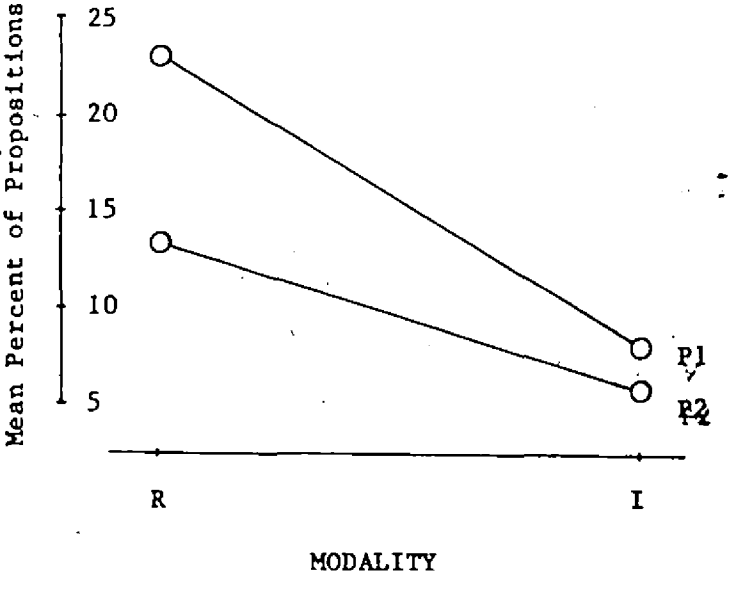


FIGURE 5: Mean Percent of Propositions Recalled or Inferred during the Q/A-with-Text Task, by Passage, Pooled over Sequence Groups. (R: Recall, I: Inference; P1: Discovery of Penicillin, P2: Magic Seeds)

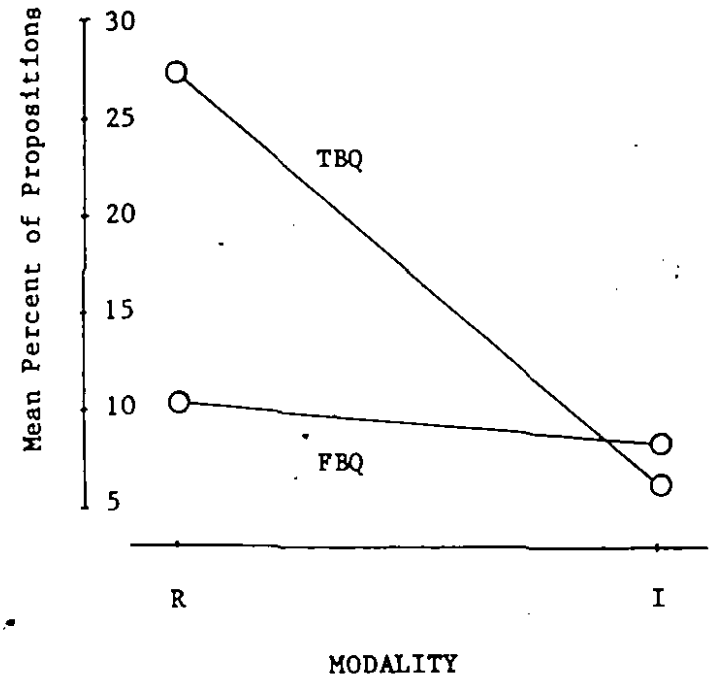


FIGURE 6: Mean Percent of Propositions Recalled or Inferred during the Q/A-with-Text Task, by Question Type, Pooled over Sequence Groups (R: Recall, I: Inference; FBQ: Frame-based Questions, TBQ: Text-based Questions)

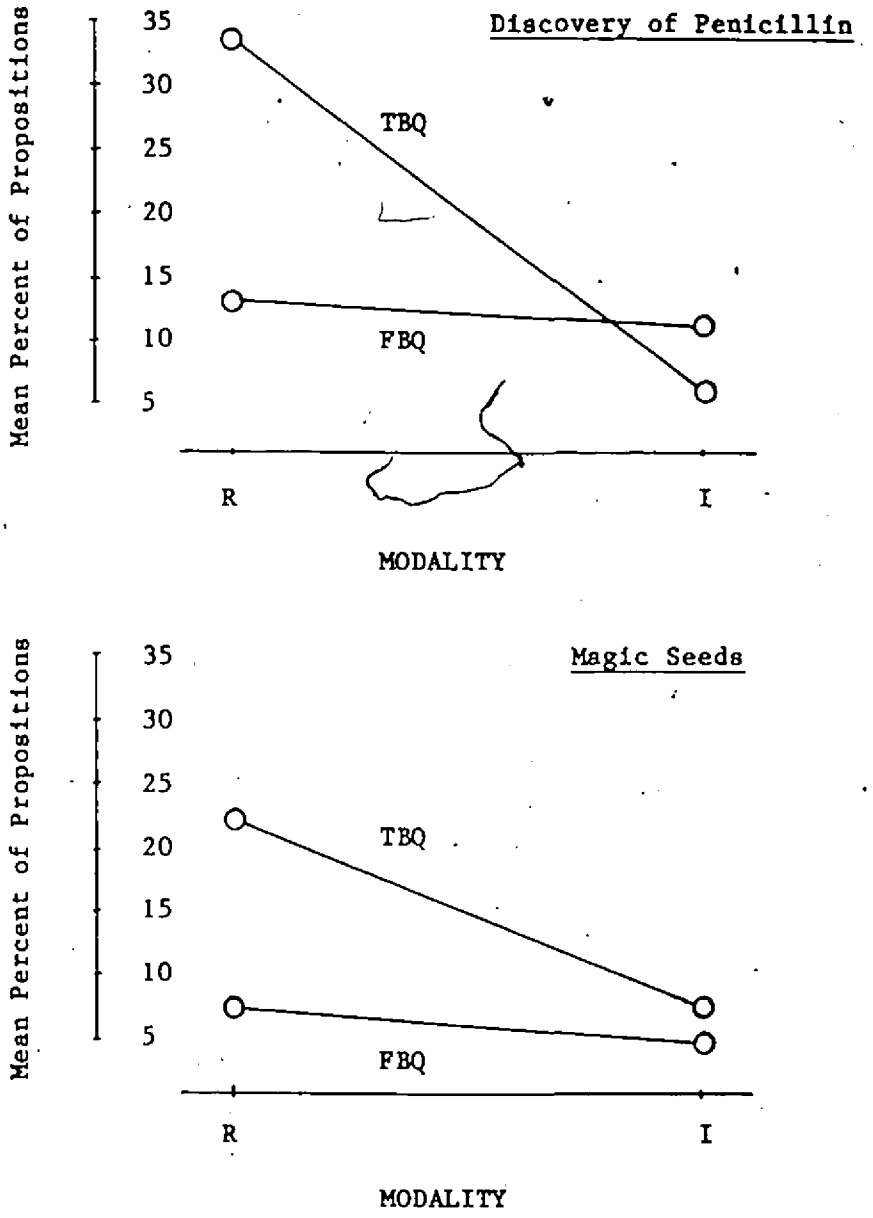


FIGURE 7: Mean Percent of Propositions Recalled or Inferred during the Q/A-with-Text Task, by Question Type and each Passage, Pooled over Sequence Groups (R: Recall, I: Inference; FBQ: Frame-based Questions, TBQ: Text-based Questions)

suggests that the differences which exist between the two types of questions in terms of the response mode that they elicit are larger for Discovery of Penicillin than for Magic Seeds.

## Analysis of the Relationship between Free Recall and Question Answering With Text

As in the case of the Q/A-no-text data, regression techniques were used to investigate the extent to which question answering with text can be accounted for by free recall. The multiple R's associated with the dependent variables are reported in Table 21, and Table 22 presents those related to the effects as derived dependent variables. For comparison, these tables also include the results obtained when free recall scores were regressed on the Q/A-no-text data (repetitions of Tables 14 and 15).

Results suggest that, once again, free recall performance is a significant predictor of the total scores students obtained while answering questions in the presence of text ($p<.0384$). Again, with one exception, the association between the two tasks is positive. Free recall scores also predict significantly the interaction between Modality of Response and Question Type ($p<.0051$). The effect of Passage is also, to some extent ($p<.0902$), accounted for by free recall data. However, only one of the scores students obtained during this task is predicted from free recall performance, and this with marginal significance ($p<.0852$) (see Table 21).

To further determine the predictability of data obtained during the Q/A-with-text task from those obtained during free

TABLE 21

Regression Analysis of Percentages of Propositions Recalled or Inferred during the Q/A Tasks with Free Recall Task Data as Predictors[1]

| DEPENDENT VARIABLES | MULTIPLE R | | F(4,14) | | p | |
|---|---|---|---|---|---|---|
| | Q/A no Text | Q/A with Text | Q/A no Text | Q/A with Text | Q/A no Text | Q/A with Text |
| P1-FBQ-R[2] | .351 | .228 | 0.493 | 0.193 | <.7412 | <.9384 |
| P1-FBQ-I | .512 | .624 | 1.241 | 2.236 | <.3386 | <.1175 |
| P1-TBQ-R[3] | .728 | .618 | 3.951 | 2.166 | <.0238* | <.1262 |
| P1-TBQ-I | .310 | .502 | 0.372 | 1.180 | <.8251 | <.3617 |
| P2-FBQ-R | .782 | .416 | 5.512 | 0.732 | <.0071** | <.5851 |
| P2-FBQ-I | .524 | .440 | 1.323 | 0.838 | <.3096 | <.5237 |
| P2-TBQ-R | .732 | .563 | 4.036 | 1.620 | <.0222* | <.2243 |
| P2-TBQ-I | .671 | .650 | 2.872 | 2.556 | <.0627*** | <.0852[a] |

1 Four predictors: % of propositions recalled (R) or inferred (I) during the free recall task for two passages (P1: Discovery of Penicillin, P2: Magic Seeds).

2 FBQ: Frame-based questions.

3 TBQ: Text-based questions.

a Standardized weights: -1.058 with P1-R as predictor; .770 with P1-I as predictor; -.164 with P2-R as predictor; .098 with P2-I as predictor.

* Significant at .05 level.

** Significant at .01 level.

*** Marginally significant.

TABLE 22

Regression Analysis of the Factors involved in the Q/A Tasks with Free Recall Task Data as Predictors[1]

| EFFECTS AS DEPENDENT VARIABLES | MULTIPLE R | | F(4,14) | | p | |
|---|---|---|---|---|---|---|
| | Q/A no Text | Q/A with Text | Q/A no Text | Q/A with Text | Q/A no Text | Q/A with Text |
| Total | .745 | .702 | 4.369 | 3.402 | <.0169* | <.0384[a] |
| Passage effect (P) | .438 | .645 | 0.831 | 2.498 | <.5272 | <.0902[b] |
| Question Type effect (Q) | .639 | .468 | 2.414 | 0.982 | <.0982** | <.4487 |
| Modality effect (M) | .435 | .431 | 0.817 | 0.796 | <.5354 | <.5472 |
| PxQ | .402 | .284 | 0.674 | 0.307 | <.6211 | <.8688 |
| PxM | .532 | .208 | 1.378 | 0.158 | <.2915 | <.9562 |
| QxM | .534 | .794 | 1.393 | 5.982 | <.2867 | <.0051[c] |
| PxQxM | .496 | .389 | 1.143 | 0.622 | <.3766 | <.6542 |

1 Four predictors: % of propositions recalled (R) or inferred (I) during the free recall task for two passages (P1: Discovery of Penicillin, P2: Magic Seeds).

a Standardized weights: .195 with P1-R as predictor; -.764 with P1-I as predictor; .186 with P2-R as predictor; .610 with P2-I as predictor.

b Standardized weights: .628 with P1-R as predictor; -.597 with P1-I as predictor; -.281 with P2-R as predictor; .558 with P2-I as predictor.

c Standardized weights: -.895 with P1-R as predictor; 1.185 with P1-I as predictor; -.742 with P2-R as predictor; .156 with P2-I as predictor.

* Significant at .05 level.

** Marginally significant.

recall, repeated-measures analysis of covariance techniques were used. The right-hand columns of Table 23 summarize the results of the ANCOVA involving the free recall data as covariates and Table 24 shows the means prior to the introduction of the covariates and the means adjusted for the effects of the covariates. The left-hand part of Table 23 repeats the results that were reported earlier in Table 20; one can thus compare the results prior to and after the introduction of the covariates. This comparison suggests that once the effects of the recall data are eliminated from the Q/A-with-text scores, the previously significant effect of Passage, its interaction with Modality and its interaction with both Modality and Question Type become non-significant. However, Modality of Response, Question Type and their interaction remain significant. Removing the variability of the free recall data from the Q/A-with-text scores has, therefore, not eliminated all significant effects and interactions.

## SUMMARY OF PRINCIPAL RESULTS

### Global Analyses of the Three Tasks

When data obtained from each of the three tasks were subjected to analyses of variance, a number of consistent results were found. Throughout all analyses, the control variable of Question Type Sequence had no significant effect, and did not significantly interact with the other factors. Consistently also, significant Passage effects were found: Students performed better in response to Discovery of Penicillin than to Magic Seeds.

TABLE 23

Repeated-Measures Analysis of Variance and Covariance of Percentages of Propositions Recalled or Inferred during the Q/A-with-Text Task

| | WITHOUT COVARIATES | | WITH COVARIATES[1] | |
|---|---|---|---|---|
| SOURCE | F(1,18) | p | F(1,14) | p |
| Between Ss: | | | | |
| Sequence of Question Type (S) | 0.236 | <.6329 | 1.251 | <.2822 |
| Within Ss: | | | | |
| Passage (P) | 32.231 | <.0001** | 1.048 | <.3234 |
| Modality of Response (M) | 78.364 | <.0001** | 7.504 | <.0160* |
| Question Type (Q) | 104.755 | <.0001** | 10.650 | <.0057** |
| PxM | 8.485 | <.0093** | 0.862 | <.3689 |
| PxQ | 0.086 | <.7732 | 0.062 | <.8071 |
| MxQ | 72.342 | <.0001** | 17.197 | <.0010** |
| PxMxQ | 18.505 | <.0005** | 1.215 | <.2889 |
| Interactions of Between and Within Ss Sources: | | | | |
| SxP | 0.092 | <.7656 | 1.508 | <.2398 |
| SxM | 1.471 | <.2409 | 0.989 | <.3369 |
| SxQ | 0.096 | <.7601 | 0.187 | <.6719 |
| SxPxM | 0.267 | <.6120 | 0.236 | <.6346 |
| SxPxQ | 0.214 | <.6490 | 0.033 | <.8579 |
| SxMxQ | 0.034 | <.8562 | 1.529 | <.2366 |
| SxPxMxQ | 0.712 | <.4100 | 0.059 | <.8120 |

1 Four covariates: % of propositions recalled or inferred during the free recall task for two passages.

* Significant at .05 level.

** Significant at .01 level.

TABLE 24

Unadjusted and Adjusted[1]
Mean Percent of Propositions Recalled or Inferred
for two Passages and two Question Types
during the Q/A-with-Text Task

| SEQUENCE GROUPS[2] | QUESTION TYPE: Frame-based | | | | QUESTION TYPE: Text-based | | | |
|---|---|---|---|---|---|---|---|---|
| | MODALITY | | | | MODALITY | | | |
| | Recall | | Inference | | Recall | | Inference | |
| | WITHOUT COVARIATES | WITH COVARIATES | WITHOUT COVARIATES | WITH COVARIATES | WITHOUT COVARIATES | WITH COVARIATES | WITHOUT COVARIATES | WITH COVARIATES |
| **Discovery of Penicillin** | | | | | | | | |
| 1 | 11.628 | 12.084 | 13.023 | 13.788 | 34.000 | 35.126 | 6.500 | 7.323 |
| 2 | 13.721 | 13.265 | 8.604 | 7.839 | 34.500 | 33.374 | 4.500 | 3.677 |
| **Magic Seeds** | | | | | | | | |
| 1 | 7.065 | 6.434 | 5.066 | 5.592 | 19.998 | 20.639 | 8.636 | 7.688 |
| 2 | 7.690 | 8.321 | 5.127 | 4.601 | 20.455 | 19.814 | 6.136 | 7.084 |

1 Means adjusted for the effects of four covariates: % of propositions recalled or inferred during the free recall task for two passages.

2 n=10 subjects per group.
Group 1 refers to children who received frame-based questions first.
Group 2 refers to those who received text-based questions first.

Another consistent finding is the difficulty of frame-related questions as compared to text-based items. This difference was, moreover, not passage-specific: In both of the Q/A tasks, the interaction of Passage by Question Type was non-significant. However, Question Type interacted significantly with Modality of Response. Text-based questions elicited answers that were more literal than inferential. In response to frame-related items on the other hand, while more inferential than factual answers were produced when children had no access to the reading material, this difference did not occur when the text was available to them. There were however significant Passage by Modality by Question Type triple interactions. In both of the Q/A tasks, the difference of the two types of question in terms of the response mode that they elicit was stronger for Discovery of Penicillin than for Magic Seeds, and, when children had no access to text, their responses to text-based questions relevant to Magic Seeds were not more literal.

Different results were however obtained in the three tasks with respect to Modality of Response (recall versus inference). The effect of Modality was significant for both the free recall and the Q/A-with-text tasks. However, while recall protocols were significantly more inferential than factual, responses given with access to text were more literal than inferential. In the case of the Q/A-no-text task on the other hand, while the Modality difference was in the same direction as that observed for free recall, it did not reach statistical significance.

Similar to the pattern of Modality-related results obtained

across tasks, is the pattern of the Modality by Passage interactions. While this interaction was non-significant for the Q/A-no-text task, it reached significance in the case of free recall and Q/A-with-text. Once again however, the two results were in opposite directions: While, in free recall, the two passages primarily differ in terms of the inferential responses that they elicit (recall protocols related to Discovery of Penicillin were more inferential than those related to Magic Seeds), in Q/A-with-text, the difference between the passages concerns the factual responses (responses related to Discovery of Penicillin were more factual than those related to Magic Seeds).

## Analyses of the Relationship Between Free Recall and the Question Answering Tasks

When data obtained from the Q/A tasks were regressed on scores on the free recall task, recall performance was found to be a significant predictor of the total scores students obtained in both of these tasks. In the case of the Q/A-no-text task, three of the eight scores students obtained were significantly accounted for by free recall data, and the effect of Question Type was also, to some extent, predicted. In the case of Q/A-with-text answers, free recall performance was a significant predictor of the interaction of Question Type by Modality. For this task, some prediction was also obtained for the effect of Passage.

When analysis of covariance procedures were applied to the Q/A data with free recall scores as covariates, all of the

previously significant effects and interactions related to Q/A-no-text data reached probabilities that were below significance level. In the case of the Q/A-with-text task however, this result was observed only in the case of the effect of Passage, its interaction with Modality, and its interaction with both Modality and Question Type. Modality, Question Type, and their interaction remained significant.

# CHAPTER VI

# DISCUSSION, CONCLUSIONS AND IMPLICATIONS

## DISCUSSION

The research reported here responds to some of the issues that were initially raised and provides rather clear-cut evidence for a number of interesting phenomena. Analyses of variance tell us about the effects of text structure and question type on comprehension and question answering, and regression and covariance analyses provide information about the relation between comprehension as evidenced by free recall on the one hand, and Q/A on the other. A number of these results should be highlighted.

### Effects of Text Structure and Question Type on Recall and Inference in Comprehension and Question Answering

Students responses were, in all tasks, unaffected by the order in which the two types of questions were presented to them. This finding makes things look less complex and more interpretable, and suggests that a theory of question answering can be, to some extent, context-free. Moreover, despite the rather small sample size (N=20), main effects and interactions were either clearly non-significant or had very low probability values. This increases our confidence that the results are most probably not limited to the specific sample used in the experiment.

Analyses of data obtained during the three discourse processing tasks provide consistent evidence for a difference between the two passages. Different types of text structure seem to elicit different processes. Magic Seeds which has a descriptive frame type (Frederiksen, 1982, in prep.; C. H. Frederiksen & J. D. Frederiksen, 1981) was more difficult than Discovery of Penicillin which is primarily narrative (op. cit.). This finding is not surprising in view of previous research (Frederiksen, 1982, in prep.) showing that narrative structures are easier for children than descriptive frames. As mentioned previously however, the narrative passage implicitly signals the use of a problem frame too. The two texts therefore differ not only with respect to type of frame but also with respect to the degree to which they directly instantiate discourse structures (see Bracewell, C. H. Frederiksen & J. D. Frederiksen, 1982). To avoid ambiguity therefore, the Passage effect is interpreted in the following way: It is highly probable that grade-six students can process language according to narrative frames more easily, even if problems and plan structures are involved in the event sequence but are not explicitly signalled; they seem, however, to be still in the process of constructing frames for processing descriptive information and even making the surface structure of the text available to them does not eliminate the relative difficulty of the descriptive passage.

Another consistent finding is the difference between the two types of questions: Whether the text is available to students or not, frame-related questions are more difficult than

text-based items. This finding supports the classification of questions. More important however is that there seem to exist distinct abilities for answering questions of different types. This is an indirect evidence for a multi-component, rather than a lump type, notion of Q/A. It must be further noted that the difference between the two question types is not passage-specific. This suggests that a theory of Q/A can be, to some extent, independent of text, and it provides higher degree of generality to the above conclusions.

The type of question asked interacted moreover with the type of response produced. This finding further supports the assumption that the two types of questions differ in the demands that they impose on the answerer's processes. As Rosch and Lloyd (1978) have put it, "answers depend on the questions asked. Unasked questions will remain unanswered. And the nature of a question constrains the kinds of answers that can be derived" (p. 1) [25]. In both of the Q/A tasks, text-based questions favoured factual responses more than inferential ones. In fact, these questions were easier for the students and it is not surprising that they responded to them more accurately. There was one exception however to this pattern: Text-based questions relevant to the descriptive passage that were answered without access to the text were not more factual than inferential (see Figure 4). This suggests once again the weakness of the cognitive representation that children of this age can construct

---

25 There is even empirical evidence showing that questions' surface structure can affect the form of answers (Levelt & Kelter, 1982).

for a descriptive text like Magic Seeds. Frame-based questions, on the other hand, in the absence of text, elicited more inferencing than recall. Once again this result can be understood in view of the difficulty that students faced in answering these questions. When the text was present during Q/A however, while frame-related questions were still answered less factually than text-based questions, the answers were not more inferential. Considering that children could look back at the original material, and that therefore this task was considerably less dependent on information storage and retrieval than earlier ones, this finding could have been expected. In summary, text-based questions tend to be associated with literal recall processes, frame-related questions elicit more inferencing, these processing differences interact with type of passage structure, and availability of reading material during Q/A creates a bias toward literal answers.

During the free recall task, children's responses were significantly more inferential than factual. This finding is typical in discourse processing research. However when students were answering questions without having access to the reading material, there was no general Modality effect. Thus, there was no significant difference between the amount of their inferential answers and that of their literal responses. It may be the case that the arbitrary distribution of the two types of questions (half of the questions were text-based and half were frame-related) has controlled the natural distribution of recall and inference. This explanation seems plausible in view of the

different modes of responses that the two types of questions elicited. The same distribution of question types was used in the Q/A-with-text task; yet, in this case, there was a general Modality effect. This suggests that this task does involve comprehension-type processes. In fact, the pattern of results obtained for this task had similarities to that of free recall in that both tasks, unlike the Q/A-no-text task, involved general Modality effects and Modality by Passage interactions. It must be noted however that, contrary to recall protocols and not surprisingly, Q/A in the presence of the reading material elicited more literal responses than inferential ones; searching for information that is literal was easier than searching for information that requires inference. Yet, it is interesting that even when they had access to the text, students did provide inferential answers.

As mentioned above, except in the case of the Q/A-no-text task -and once again the mentioned question-type-distribution explanation seems plausible-, the difference between recall and inference was passage-specific thus suggesting that frame differences are associated with processing differences. Cognitive adaptation to the demands of different discourse structures seems to have occured. In the recall task, the superiority of inferencing over literal recall is much more pronounced for Discovery of Penicillin. Free recall performance is not only higher for the narrative, but is also much more inferential. Children seem to have built a more accurate representation for the narrative and they manipulated it more

actively. The pattern of results in the last task, on the other hand, is almost exactly the opposite of that of the recall task (compare Figures 2 and 5). Performance is still higher for the narrative, but, the surface structure of the material being available, responses are more factual especially in the case of the easier passage. The students seem to lack the ability to create a frame for the descriptive text, and therefore difficulty in searching for information that requires inference is heightened.

## The Relationship Between Comprehension and Question Answering

When the predictability of Q/A data from free recall scores was investigated by regression techniques, it was found that the general level of performance in both of the Q/A tasks depended significantly on performance in recall. Moreover, the results of the analyses of covariance suggest some overlap between comprehension and question answering. The pattern of results however was not identical for Q/A with and without access to text. Results obtained in the case of each Q/A task are discussed next.

### The Relationship Between Comprehension and Question Answering Without Access to Text

As mentioned above, the general level of performance in answering questions without access to text was found predictable from comprehension as evidenced by free recall performance. It seems that the ability to answer questions in general depends on the availability of information in memory and that the extent of

memory structure for text is a determinant of performance in Q/A. There was also some prediction of the Question Type effect. Considering the processing differences that were found between the two types of questions, it seems that the ability to answer text-based questions is primarily dependent on recalled propositions and that performance in frame-based items is related to structures that are inferred in comprehension. There is moreover some dependency between performance in questions and the structure of text.

The results of the analysis of covariance suggest that, despite the question-induced demands involved in the process of responding to questions, there seems to exist considerable similarity between Q/A and comprehension as evidenced by free recall. As discussed in the Literature Review and the Rationale sections, Q/A and comprehension cannot logically be underlied by identical cognitive processes. However, when, using the free recall data as covariates, individual differences were 'controlled', all the factors that were significantly affecting the Q/A-no-text data became non-significant. Thus, the significance of the effects of Passage and Question Type was removed when individual differences in memory structure were taken into account. Level of performance in Q/A, especially in answering questions relevant to the descriptive passage, reflects level of free recall, and, since availability of frame information is important to answering frame-based questions, the result concerning the effect of Question Type suggests that comprehension involves frame-related aspects.

The similarity found between comprehension and Q/A may be due to shared processes or shared cognitive structures. In view of the specific pattern of results, and considering that, contrary to the free recall and the Q/A-with-text tasks, there was no significant Passage by Modality interaction in the Q/A-no-text data and that the effect of Passage was removed in the analysis of covariance, the shared-structures explanation seems more probable. Obviously, we cannot rule out the process-based hypothesis, but the results seem to be more consistent with the notion that Q/A performance depends on available memory structures and that question effects reflect memory structure differences. Specific inferential processes required by different types of questions are most likely to occur if relevant propositional or inferred frame structures are present in memory.

Presented with questions, students seem to have constructed their answers essentially by accessing cognitive structures already present in memory, rather than constructing new schemes on the basis of the stored information. To a large extent therefore, common structures underlie comprehension and question answering.

The Relationship Between Comprehension and Question Answering in the Presence of Text

Similar to question answering without access to the original material, the general level of performance in Q/A in the presence of text was found predictable from comprehension. Free recall task variables also accounted for the interaction of

Question Type by Modality of Response. There was moreover some predictability of Passage difference in Q/A-with-text from the comprehension data. Here again both a memory-based model and a shared-process explanation are possible. However, in view of the fact that, in this task, the students had access to the text, a memory-based explanation is very unlikely. Rather, it seems that the predictability of performance in this task from comprehension is due to one's general processing efficiency in handling the common processes that underlie the two tasks.

When, in the analysis of covariance, individual differences in the recall task were considered, the previously significant effect of Passage and its interaction with Modality of Response obtained chance level probabilities. It seems therefore that comprehension-type processes are involved in Q/A-with-text. Moreover, the statistical significance of the triple interaction of Passage by Modality by Question Type was removed when individual scores were adjusted for the covariates, i.e., the free recall variables. This suggests that text- and frame-based processes are involved in comprehension. In view of the fact that all factors that included Passage became non-significant in the analysis of covariance, and as mentioned earlier, considering that the text was available to students, a structure-oriented interpretation of these results does not seems plausible. There appears to be a partial overlap in terms of the processes that are involved in comprehension and Q/A-with-text. Hence, even in presence of the text one does have to understand it in order to answer questions relevant to it.

However, the results of the analysis of covariance suggest that answering questions while looking back at the original material also involves processes distinct from those involved in comprehension. In fact, when data obtained from the recall task were regressed on scores of the with-text answers, the residuals were still significantly affected by Question Type, Modality of Response, and their interaction. It is reasonable to assume that factors that remained significant reflect at least partially the ability to perform search operations[26] and to locate the queried information. Those previously significant effects and interactions that did become non-significant with the introduction of the covariates, on the other hand, must reflect common processes. These however do not account for the Q/A-with-text data, suggesting that this task involves unique demands. The ability to answer questions with the text available for reference most probably involves processes independent from the ability to comprehend it.

## CONCLUSIONS, AND THEORETICAL AND EDUCATIONAL IMPLICATIONS

This research has investigated the relationship between comprehension and the mechanisms that underlie question answering, and has examined the effects of discourse structure

---

26 Search operations must also be involved when answers are produced without access to text. However, while Q/A-with-text involves searching the actual text, search strategies in Q/A-no-text tasks operate on the mental representation of the text.

and question type on recall and inference in these tasks. It has provided empirical evidence which shows that different types of questions elicit different processes that are text-structure dependent. It has moreover provided data consistent with the notion that, globally, comprehension as reflected in recall protocols and answering questions after reading a text without referring back to it do not differ considerably. The cognitive demands that are induced by questions do not seem to lead comprehenders to construct new structures, and question answering in the absence of text seems to reflect principally the availability of question-relevant information in memory representation for text. Assuming that the questions used in the study were genuine text- and frame-related ones, we can conclude that both text-based, bottom-up and framing, top-down processes were occuring during comprehension and recall. Considering the relative difficulty of frame-based questions however, subjects' comprehension seems to have been oriented more towards bottom-up processes.

As mentioned in the review of the literature on Q/A, several researchers have emphasized the differences between comprehension and Q/A, and a number of them (Anderson & Bower, 1974; Lindsay & Norman, 1972) have drawn an analogy between Q/A and problem solving. The results of the present study are not incompatible with these theories of Q/A. Q/A and problem solving may be similar and it is certainly true that the processes of Q/A and comprehension are not identical. However, it may be the

case that research on comprehension has underestimated its active, constructive nature[27].

Parallel to the question-answering/problem-solving analogy, is the analogy drawn by Collins, Brown, and Larkin (1980) and Frederiksen (1981, 1982, in prep.) between comprehension and problem solving[28]. It can be reasonably hypothesized that comprehension, question answering, problem solving, and, as Greeno (1974) has suggested, learning all involve processes of constructing conceptual structures in memory, and are therefore similar in that respect.

There is no need for empirical support for the logically obvious notion that Q/A and comprehension are not identical. Any cognitive theory of Q/A which strives for a testable level of explicit detail would have to specify the differences between Q/A and comprehension. However, relinquishing to some extent the press for explicitness that characterizes information-processing psychology (Floden, 1981)[29], the results of this study suggest

---

27 This issue is related to the criticisms addressed to excessively text- or schema-based conceptions of discourse processing, discussed in the Process section of Chapter 2.

28 While those who draw an analogy between Q/A and problem solving have a primarily means-ends-analysis conception of problem solving, Collins et al. and Frederiksen, whose work is more recent, view problem solving as a structure-building process.

29 Floden (1981) has attributed this feature of modern cognitive psychology to the influence of computer science, Winograd (1977) has pointed out the overformalization that results from the desire for "getting a [computer] system to work" (p. 85), and J. R. Anderson, Kline, and Lewis (1977) have pointed out that with the high level of specification needed in computer simulation, "it is almost certain that, whatever we attempt to model, the model will be wrong [...] but [...]" (p. 309).

that, at a global level, if someone has understood a unit of information, s/he must be able to answer comprehension questions relevant to it.

Obviously this conclusion is at least limited by the type of question asked. The questions that were used in this study were intended to be 'straight comprehension' questions. Conceivably, some other types of questions can force the language user to draw inferences and construct structures other than those drawn and built spontaneously during comprehension. In fact, Kubes (1982) has shown that comprehension is not sufficient for integration, i.e., for answering questions requiring the connection of information acquired in different passages.

It remains also to be seen whether there are significant qualitative differences between information produced in free recall and that produced in response to questions. More fine-grained, primarily qualitative analyses are needed to determine the nature of the relationship between comprehension and Q/A. Frame analysis of discourse (Frederiksen, 1982, in prep.) provides a basis for a principled way of conducting such analyses. Also, analyses involving the classification of the inferences (Frederiksen, 1979, 1981) drawn in each task could tell us more about how comprehension and Q/A are related. Further research is needed to determine whether the similarities found between comprehension and question answering reflect primarily shared processes or shared cognitive structures.

These limitations also apply to the results regarding the

relationship between comprehension and Q/A in the presence of text. The study provides evidence showing that answering questions while referring to the text involves both comprehension-type processes and specific question answering abilities. It thus reflects comprehension as well as independent Q/A processes. Moreover, it biases the language user towards a more literal representation of discourse than one normally constructs.

What were the students doing cognitively during this task however? Frase (1975) has argued that memory is affected by the re-encoding of information. Re-encoding is an important feature of this task, but at this point, we can only speculate on the processes that underlie the task. It is highly probable that, as Nicholson and Imlach (1981) and Johnston (1982) have argued, students were simply searching for relevant information and manipulated it minimally to generate answers. It is also possible that they were using the text to fill the missing information of the structures that they had built previously. Another possibility is that they were constructing frames that they had failed to construct in the absence of the text, under surface structure constraints. There may also be considerable individual differences and contextual variation in what students choose to do in such a task. For an understanding of the actual cognitive activities of students during this task, more detailed, mainly qualitative analyses are needed (e.g., coding students responses on the basis of the frame analysis of the passages and comparing them to a frame-based coding of

propositions produced in the absence of text). It would also be interesting to conduct a study in which Q/A with text is not preceded by reading, recall, and Q/A without text.

In any case, the pattern of relationships that was found between comprehension and the two Q/A tasks has a number of educational implications. In a Q/A task where the text is available to students, comprehension processes are involved but so are question-specific processes of information search and inference. What such a task measures depends on the type of questions asked and it also produces highly literal responses. If, on the other hand, questions are answered after reading but without access to the text, responses would principally reflect the memory structures that were generated during comprehension. Thus, while Q/A-no-text more directly reflects previous comprehension, the frequently-used task of Q/A in the presence of text involves comprehension and other processes and does not seem an adequate reflection of what comprehenders do spontaneously to make sense of a text. As Johnston (1982) has suggested therefore, it is not an appropriate method for the assessment of comprehension.

The implications of another consistent finding of this study are also to be noted. It was found that the descriptive passage was more difficult for students than the narrative. It does not seem however that textbook writers are informed of such findings and, without wishing to generalize from a small sample, it is interesting to note that the grade-six teachers who examined the two passages of this study commented that Magic

Seeds, i.e., the descriptive text, would be the easier of the two for the children. Moreover, in one of the passages, namely, Discovery of Penicillin, some important structural aspects, i.e., the problem frame and the plan structure, are rarely stated explicitly. The text seems to disguise these aspects, presumably to make it more comprehensible for children. Considering the relative difficulty of (problem) frame-related questions over text-based items, making the text more and more concrete and only implying the high-level goals that direct the event sequence do not seem to have helped the students. In view of their responses to frame-related questions, it seems that children have difficulty understanding text structures that we may expect them to understand.

Further research is needed to ensure the generalizability of the above conclusions and studies of a more micro-analytic, process-oriented nature are necessary for a fine-grained characterization of the processes involved in understanding discourse and answering different types of questions relevant to various text structures.

In view of the importance of reading comprehension and question answering in the information processes of our society in general and education in particular, and considering that they are, for obvious reasons, of special interest to cognitive scientists, much more is to be known about comprehension and question answering. This study suggests that the former may be just as constructive as the latter.

## REFERENCES


Abelson, R. P. Script processing in attitude formation and decisionmaking. InJ. S. Carroll, & J.W. Payne(Eds.), Cognition and social behavior. Hillsdale, N.J: Erlbaum, 1976.

Adams, M. J., & Collins, A. A schema-theoretic view of reading. In R. O. Freedle (Ed.), New directions in discourse processing. Norwood, N.J.: Ablex, 1979.

Anderson, J. R., & Bower G. H. Human associative memory. Washington, D.C.: Hemisphere, 1974.

Anderson, J. R., Kline, P.J., & Lewis C. H. A production system model of language processing. In M. A. Just & P. A. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, N.J.: Erlbaum, 1977.

Anderson, R. C. How to construct achievement tests to assess comprehension. Review of Educational Research, 1972, 42, 145-170.

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

Anderson, R. C., & Ortony A. On putting apples into bottles- A problem of polysemy. Cognitive Psychology, 1975, 7, 167-180.

Anderson, R. C., Spiro, R. J., & Anderson, M. C. Schemata as scaffolding for the representation of information in connected discourse. American Educational Research Journal, 1978, 15, 433-440.

Anderson, T. H. Study strategies and adjunct aids. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension. Hillsdale, N. J.: Erlbaum, 1980.

Anderson, T. H., Wardrop, J. L., Hively, W., Muller, K., Anderson, R. I., Hastings, N. C., & Frederiksen, J. Plans and procedures for developing and using domain-referenced tests. Urbana, Ill.: Center for the Study of Reading, University of Illinois and the B.B.N., 1977.

Andre, T. Does answering higher-level questions while reading facilitate productive learning? Review of Educational Research, 1979, 49, 280-318.

Andre, T. The role of paraphrased adjunct questions in facilitating learning from prose. Contemporary Educational Psychology, 1981, 6, 22-27.

Andre, T., Mueller, C., Womack, S., Smid, K., & Tuttle, M. Adjunct application questions facilitate later application, or do they? Journal of Educational Psychology, 1980, 72, 533-543.

Andre, T., & Sola, J. Imagery, verbatim and paraphrased questions, and retention of meaningful sentences. Journal of Educational Psychology, 1976, 68, 661-669.

Ausubel, D. P. The use of advance organizers in the learning and retention of meaningful verbal material. Journal of Educational Psychology, 1960, 51, 267-272.

Ausubel, D. P. Cognitive structure and the facilitation of meaningful verbal learning. In R. C. Anderson, & D. P. Ausubel (Eds.), Readings in the psychology of cognition. New York: Holt, Rinehart & Winston, 1965.

Ausubel, D. P., & Fitzgerald, D. Organizer, general background, and antecedent learning variables in sequential verbal learning. In R. C. Anderson, & D. P. Ausubel (Eds.), Readings in the psychology of cognition. New York: Holt, Rinehart & Winston, 1965.

Baker, E. L. Beyond objectives: Domain-referenced tests for evaluation and instructional improvement. In W. Hively (Ed.), Domain-referenced testing. Englewood Cliffs, N. J.: Educational Technology Publications, 1974.

Barnes, B. R., & Clawson, E. U. Do advance organizers facilitate learning? Recommendations for further research based on an analysis of 32 studies. Review of Educational Research, 1975, 45, 637-659.

Bartlett, F. C. Remembering: A study in experimental and social psychology. Cambridge , England: Cambridge University Press, 1932.

de Beaugrande, R. Text, discourse, and process. Toward a multidisciplinary science of texts. Norwood, N. J.: Ablex, 1980.

de Beaugrande, R. Design criteria for process models of reading. Reading Research Quarterly, 1981, 16, 261-315.

de Beaugrande, R., & Miller, G. W. Processing models for children's story comprehension. Poetics, 1980, 9, 181-201.

Bing, S. B. Role of adjunct questions and reading ability levels on rote and conceptual learning from prose. Instructional Science, 1982, 11, 129-138.

Black, J. B., & Bower, G. H. Story understanding as problem-solving. Poetics, 1980, 9, 223-250.

Bloom, B. S. (Ed.), Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York: David McKay Company, Inc., 1956.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill, 1971.

Bobrow, D. G., & Norman, D. A. Some principles of memory schemata. In D. G. Bobrow, & A. Collins (Eds.), Representation and understanding: Studies in cognitive science. New York: Academic Press, 1975.

Bormuth, J. R. On the theory of achievement test items, Chicago: The University of Chicago Press, 1970.

Bormuth, J. R., Manning, J., Carr, J., & Pearson, D. Children's comprehension of between- and within-sentence syntactic structures. Journal of Educational Psychology, 1970, 61, 349-357.

Bracewell, R. J., Frederiksen, C. H., & Frederiksen, J. D. Cognitive processes in composing and comprehending discourse. Educational Psychologist, 1982, 17, 146-164.

Bransford, J. D., Barclay, R. J., & Franks, J. J. Sentence memory: A constructive versus interpretive approach. Cognitive Psychology, 1972, 3, 193-209.

Bransford, J. D., & Franks, J. J. The abstraction of linguistic ideas. Cognitive Psychology, 1971, 2, 331-350.

Bruce, B. C. Analysis of interacting plans as a guide to the understanding of story structure. Poetics, 1980, 9, 295-311.(a)

Bruce, B. C. Plans and social action. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension. Hillsdale, N.J.: Erlbaum, 1980.(b)

Bruning, R. H. Effects of review and testlike events within the learning of prose material. Journal of Educational Psychology, 1968, 59, 16-19.

Carrier, C. A., & Fautsch-Patridge, T. Levels of questions: A framework for the exploration of processing activities. Contemporary Educational Psychology, 1981, 6, 365-382.

Chafe, W. L. Creativity in verbalization and its implications for the nature of stored knowledge. In R. O. Freedle (Ed.),

Discourse production and comprehension. Norwood, N.J.: Ablex, 1977.

Clark, H. H. Inferences in comprehension. In D. LaBerge, & S. J. Samuels (Eds.), Basic processes in reading: Perception and comprehension. Hillsdale, N.J.: Erlbaum, 1977.

Clark, H. H., & Haviland, S. E. Comprehension and the given-new contract. In R. O. Freedle (Ed.), Discourse production and comprehension. Norwood, N.J.: Ablex, 1977.

Clements, P. The effects of staging on recall from prose. In R. O. Freedle (Ed.), New directions in discourse processing. Norwood, N. J.: Ablex, 1979.

Collins, A. Processes in aquiring knowledge. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, N.J.: Erlbaum, 1977.

Collins, A., Brown, A. L., Morgan, J. L., & Brewer, W. F. The analysis of reading tasks and texts. Urbana, Ill.: Center for the Study of Reading. University of Illinois and the B.B.N., 1977.

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.

Collins, A. Warnock, E. H., Aiello, N., & Miller, M. L. Reasoning from incomplete knowledge. In D. G. Bobrow & A. Collins (Eds.), Representation and understanding: Studies in cognitive science. New York: Academic Press, 1975.

Collins, A., Warnock, E. H., & Passafiume, J. J. Analysis and synthesis of tutorial dialogues. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 9). New York: Academic Press, 1975.

Cook, T. D., & Campbell, D. T. Quasi-experimentation. Chicago: Rand McNally, 1979.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 671-684.

Danks, J. H., & Glucksberg, S. Experimental psycholinguistics. Annual Review of Psychology, 1980, 31, 391-411.

Dillon, J. T. The multidisciplinary study of questioning. Journal of Educational Psychology, 1982, 74, 147-165.

Dooling, J. D., & Lachman, R. Effects of comprehension on retention of prose. Journal of Experimental Psychology, 1971, 88, 216-222.

Dore, J., Gearhart, M., & Newman, D. The structure of nursery school conversation. In K. E. Nelson (Ed.), Children's language (Vol. 1). New York: Gardner Press, 1978.

Ebbesen, E. B. Cognitive processes in inferences about a person's personality. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), Social cognition. The Ontario Symposium (Vol. 1). Hillsdale, N.J.: Erlbaum, 1981.

Ellis, J. A., Konoske, P. J., Wulfeck II, W. H., & Montague, W. E. Comparative effects of adjunct postquestions and instructions on learning from text. Journal of Educational Psychology, 1982, 6, 860-867.

Felker, D. B., & Dapra, R. A. Effects of question type and question placement on problem-solving ability from prose material. Journal of Educational Psychology, 1975, 67, 380-384.

Feurzeig, W., Munter, P., Swets, J., & Breen, M. Computer-aided teaching in medical diagnosis. Journal of Medical Education, 1964,39, 746-754.

Fillmore, C. J. The case for case. In E. Bach & R. T. Harms (Eds.), Universals in linguistic theory. New York: Holt, Rinehart & Winston, 1968.

Finn, P. J. A question writing algorithm. Journal of Reading Behavior, 1975, 7, 341-367.

Floden, R. E. The logic of information-processing psychology in education. Review of Research in Education, 1981, 9, 75-109.

Frase, L. T. Learning from prose material: Length of passage, knowledge of results, and positions of questions. Journal of Educational Psychology, 1967, 58, 266-272.

Frase, L. T. Questions as aids to reading. Some research and a theory. American Educational Research Journal, 1968, 5, 319-332.

Frase, L. T. Some data concerning the mathemagenic hypothesis. American Educational Research Journal, 1968, 5, 181-189. (b)

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

Frederiksen, C. H. Effects of task-induced cognitive operations on comprehension and memory processes. In J. B. Carroll, & R. O. Freedle (Eds.), Language comprehension and the acquisition of knowledge. Washington, D.C.: V. H. Winston, 1972.

Frederiksen, C. H. Representing logical and semantic structure of knowledge acquired from discourse. Cognitive Psychology, 1975, 7, 371-458.

Frederiksen, C. H. Semantic processing units in understanding text. In R. O. Freedle (Ed.), Discourse production and comprehension. Norwood, N.J.: Ablex, 1977.(a)

Frederiksen, C. H. Structure and process in discourse production and comprehension. In M. A. Just, & P. A. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, N.J.: Erlbaum, 1977.(b)

Frederiksen, C. H. Discourse comprehension and early reading. In L. B. Resnick, & P. A. Weaver (Eds.), Theory and practice of early reading. Hillsdale, N.J.: Erlbaum, 1979.

Frederiksen, C. H. Inference in preschool children's conversations: A cognitive perspective. In J. Green, & C. Wallat (Eds.), Ethnography and language in educational settings. Norwood, N.J.: Ablex, 1981.

Frederiksen, C. H. The relationship of frame construction and comprehension of extended school-type text. Paper presented at the American Educational Research Association meeting, New York, 1982.

Frederiksen, C. H. Knowledge and inference in discourse communication. Norwood, N.J.: Ablex.(Book in preparation).

Frederiksen, C. H., & Frederiksen, J. D. The problem of genre knowledge in comprehension: Prototype text structures. Paper presented at the American Educational Research Association meeting, Los Angeles, 1981.

Frederiksen, C.H., Frederiksen, J. D., & Bracewell, R. J. Discourse analysis of children's text production. In A. Matsuhashi (Ed.), Writing in real time. Longman, in press.

Frederiksen, C. H., Frederiksen, J. D., Humphrey, F. M., & Ottesen J. Discourse inference: Adapting to the inferential demands of school texts. Paper presented at the American Educational Research Association meeting, Toronto, 1978.

Friedman, F., & Rickards, J. P. Effect of level, review, and sequence of inserted questions on text processing. Journal of Educational Psychology, 1981, 73, 427-436.

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.

Geertz, C. Thick description: Toward an interpretive theory of culture. In C. Geertz, The interpretation of cultures. New York: Basic Books, 1973.

Glaser, R. Instructional technology and the measurement of learning outcomes: Some questions. American Psychologist, 1963, 18, 519-521.

Green, J. A. Teacher-made tests (2nd ed.). New York: Harper & Row, 1975. (Originally published, 1963.)

Greene, D. Social perception as problem solving. In J. S. Carroll, & J. W. Payne (Eds.), Cognition and social behavior. Hillsdale, N.J.: Erlbaum, 1976.

Greeno, J. G. Processes of learning and comprehension. In L. W. Gregg (Ed.), Knowledge and cognition. Potomac, Md.: Erlbaum, 1974.

Greeno, J. G. Cognitive objectives of instruction: Theory of knowledge for solving problems and answering questions. In D. Klahr (Ed.), Cognitive and instruction. Hillsdale, N.J.: Erlbaum, 1976.

Grice, H. P. Logic and conversation. In P. Cole & J. L. Morgan (Eds.), Syntax and semantics; Vol. 3: Speech acts. New York: Seminar Press, 1975.

Grimes, J. The thread of discourse. The Hague: Mouton, 1975.
Halliday, M. A. K., & Hasan, R. Cohesion in English. London: Longman, 1976.

Hamilton, D. L. Cognitive representations of persons. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), Social cognition. The Ontario Symposium (Vol. 1). Hillsdale, N.J.: Erlbaum, 1981.

Hastie, R. Schematic principles in human memory. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), Social cognition. The Ontario Symposium (Vol. 1). Hillsdale, N.J.: Erlbaum, 1981.

Hastie, R. Social inference. Annual Review of Psychology, 1983, 34, 511-542.

Hayes-Roth, B., & Thorndyke, P. W. Integration of knowledge from text. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 91-108.

Hively, W. (Ed.), Domain-referenced testing. Englewood Cliffs, N. J.: Educational Technology Publications, 1974.

Hunkins, F. P. Effects of analysis and evaluation questions on various levels of achievement. Journal of Experimental Education, 1969, 38, 45-58.

Hunt, E. B. Artificial intelligence. New York: Academic Press, 1975.

Jensen K., & Wirth N. Pascal user manual and report (2nd ed.). New York: Springer Verlage, 1974.

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

Johnston, P. Implications of basic research for the assessment of reading comprehension. Urbana, Ill.: Center for the Study of Reading, University of Illinois and the B.B.N., 1981.

Johnston, P. Question type and the assessment of reading comprehension. Paper presented at the American Educational Research Association meeting, New York, 1982.

Jones, H. E. Experimental studies of college teaching. The effect of examination on permanence of learning. Archives of Psychology, 1923, 10, 1-70.

Kaplan, S. J. Appropriate responses to inappropriate questions. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), Elements of discourse understanding. Cambridge: Cambridge University Press, 1981.

Katz, S., Atkeson, B., & Lee, J. The Bransford-Franks linear effect: Integration or artifact? Memory and Cognition, 1974, 2, 709-713.

Katz, S., & Gruenewald, P. The abstraction of linguistic ideas in "meaningless" sentences. Memory and Cognition, 1974, 2, 737-741.

Keenan, J. M., & Kintsch, W. The identification of explicitly and implicitly presented information. In W. Kintsch (Ed.), The representation of meaning in memory, Hillsdale, N.J.: Erlbaum, 1974.

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

Kintsch, W. On comprehending stories. In M. A. Just, & P. A. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, N.J.: Erlbaum, 1977.

Kintsch, W. Comprehension and memory of text. In W. K. Estes (Ed.), Handbook of learning and cognitive processes (Vol. 6). Hillsdale, N.J.: Erlbaum, 1978.

Kintsch, W., Kozminsky, E., Streby, W. J., McKoon, G., & Keenan, J. M. Comprehension and recall of text as a function of content variables. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 196-214.

Kintsch W., & van Dijk, T. A. Comment on se rappelle et on résume

des histoires. Langages, 1975, 40, 98-116.

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

Kohn, B. Our tiny servants. Prentice Hall, Inc., 1962.

Kormos, L. Individual differences in children's discourse comprehension under various questioning conditions. Paper presented at the American Educational Research Association meeting, Montreal, 1983.

Kubes, M. Integrative processes in the acquisition of knowledge from text. Unpublished master's thesis, McGill University, 1982.

Kuhn, T. S. The structure of scientific revolutions (2nd ed.). Chicago: University of Chicago Press, 1970.

Lachman, J. L., & Lachman, R. Comprehension and cognition: A state of the art inquiry. In L. Cermak, & F. I. M. Craik (Eds.), Levels of processing in human memory. Hillsdale, N.J.: Erlbaum, 1979.

Lehnert, W. G. The process of question answering. A computer simulation of cognition. Hillsdale, N.J.: Erlbaum, 1978.

Lehnert, W. G. A computational theory of human question answering. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), Elements of discourse understanding. Cambridge: Cambridge University Press, 1981.

Levelt, W. J. M. & Kelter, S. Surface form and memory in question answering. Cognitive Psychology, 1982, 14, 78-106.

Lindsay, P. H., & Norman D. A. Human information processing. New York: Academic Press, 1972.

Lindvall, C. M., & Nitko, A. J. Measuring pupil achievement and aptitude (2nd ed.). New York: Harcourt Brace Jovanovich, Inc., 1975.

Lucas, P. A., & McConkie, G. W. The definition of test items: A descriptive approach. American Educational Research Journal, 1980, 17, 133-140.

Mager, R. F. Preparing instructional objectives. Palo Alto, California: Fearon Publishers, 1962.

Magoon, A. J. Constructivist approaches in educational research. Review of Educational Research, 1977, 47, 651-693.

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

111-151.

Marshall, N., & Glock, M. D. Comprehension of connected discourse: A study into the relationship between the structure of text and information recalled. Reading Research Quarterly, 1978, 14, 10-56.

McConkie, G. W. Learning from text. Review of Research in Education, 1977, 5, 3-48.

McGaw, B. & Grotelueschen, A. Direction of the effect of questions in prose material. Journal of Educational Psychology, 1972, 63, 580-588.

McKoon, G., & Keenan, J. M. Response latencies to explicit and implicit statements as a function of the delay between reading and test. In W. Kintsch (Ed.), The representation of meaning in memory. Hillsdale, N.J.: Erlbaum, 1974.

Meyer, B. J. F. The structure of prose: Effects on learning and memory and implications for educational practice. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, N.J.: Erlbaum, 1977.(a)

Meyer, B. J. F. What is remembered from prose: A function of passage structure. IN R. O. Freedle (Ed.), Discourse production and comprehension. Norwood, N.J.: Ablex, 1977.(b)

Meyer, B. J. F., Haring, M. J., Brandt, D. M., & Walker, C. H. Comprehension of stories and expository text. Poetics, 1980, 9, 203-211.

Millman, J. Criterion-referenced measurement. In J. W. Popham (Ed.), Evaluation in education. Berkeley, California: McCutchan, 1974.

Minsky, M. A framework for representing knowledge. In P. H. Winston (Ed.), The psychology of computer vision. New York: McGraw Hill, 1975.

Morgan, J. L., & Green, G. M. Pragmatics and reading comprehension. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension. Hillsdale, N.J.: Erlbaum, 1980.

Morgan, J. L., & Sellner, M. B. Discourse and linguistic theory. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension. Hillsdale, N.J.: Erlbaum, 1980.

Nicholas, D. W., & Trabasso, T. Toward a taxonomy of inferences for story comprehension. In F. Wilkening, J. Becker, & T. Trabasso(Eds.), Information integration by children. Hillsdale, N.J.: Erlbaum, 1980.

Nicholson, T., & Imlach, R. Where do their answers come from? A study of the inferences which children make when answering questions about narrative stories. Journal of Reading Behavior, 1981, 13, 111-129.

Nicholson, T., Pearson, P. D., & Dykstra, R. Effects of embedded anomalies and oral reading errors on children's understanding of stories. Journal of Reading Behavior, 1979, 11, 339-354.

Norman, D. A. Memory, knowledge, and the answering of questions. In R. L. Solso (Ed.), Contemporary issues in cognitive psychology: The Loyola symposium. New York: Halsted Press, 1972.

Norman, D. A. Memory and attention (2nd ed.). New York: John Wiley, 1976.

Norman, D. A., Rumelhart, D. E., & The LNR Research Group. Explorations in cognition. San Francisco: Freeman, 1975.

Parlebas, P. Un modèle d'entretien hyper-directif: La maïeutique de Socrate. Revue Française de Pédagogie, 1980, 51, 4-19.

Pearson, P.D., & Johnson, D. D. Teaching reading comprehension. New York: Holt, Rinehart & Winston, 1978.

Reder, L. M. The role of elaborations in memory for prose. Cognitive Psychology, 1979, 11, 221-234.

Reder, L. M. The role of elaboration in the comprehension and retention of prose: A critical review. Review of Educational Research, 1980, 50, 5-53.

Resnick, L. B. Holding an instructional conversation: Comments on chapter 10 by Collins. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, N.J.: Erlbaum, 1977.

Rickards. J. P. Adjunct postquestions in text: A critical review of methods and processes. Review of Educational Research, 1979, 49, 181-196.

Rickards, J. P. Notetaking, underlining, inserted questions, and organizers in text: Research conclusions and educational implications. Educational Technology, 1980, 20, 5-11.

Rieger III, C. J. Conceptual memory and inference. In R. C. Schank, Conceptual information processing. Amsterdam: North-Holland, 1975.

Robinson, F. P. Effective study (4th ed.). New York: Harper & Row, 1970. (Originally published, 1941.)

Rosch, E., & Lloyd, B.B. (Eds.), Cognition and categorization. Hillsdale, N.J.: Erlbaum, 1978.

Rothkopf, E. Z. Some theoretical and experimental approaches to problems in written instruction. In J. D. Krumboltz (Ed.), Learning and the educational process. Chicago: Rand McNally, 1965.

Rothkopf, E. Z. Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. American Educational Research Journal, 1966, 3, 241-249.

Rothkopf, E. Z. Structural text features and the control of processes in learning from written materials. In J. B. Carroll, & R.O. Freedle (Eds.), Language comprehension and the acquisition of knowledge. Washington, D.C.: V. H. Winston, 1972.

Rothkopf, E. Z. Writing to teach and reading to learn: A perspective on the psychology of written instruction. In N. L. Gage (Ed.), The psychology of teaching methods. The Seventy-Fifth Yearbook of the National Society for the Study of Education, Part 1. Chicago, Illinois: University of Chicago Press, 1976.

Rothkopf, E. Z., & Bisbicos, E. E. Selective facilitative effects of interpersed questions on learning from written prose. Journal of Educational Psychology, 1967, 58, 56-61.

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

Rumelhart, D. E. Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension. Hillsdale, N.J.: Erlbaum, 1980.

Rumelhart, D. E., & Ortony, A. The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, N.J.: Erlbaum, 1977.

Sachs, J. S. Recognition memory for syntactic and semantic aspects of connected discourse. Perception and Psychophysics, 1967, 2, 437-442.

Sadow, M. W. The use of story grammar in the design of questions. The Reading Teacher, 1982, 35, 518-522.

Schank, R. C. Conceptual dependency: A theory of natural language understanding. Cognitive Psychology, 1972, 3, 552-631.

Schank, R. C., & Abelson, R. P. Scripts, plans, goals and understanding: An inquiry into human knowledge structures. Hillsdale, N.J.: Erlbaum, 1977.

Schank, R. C., & Lebowitz, M. Levels of understanding in computers and people. Poetics, 1980, 9, 251-273.

Shavelson, R. J., Berliner, D. C., Ravitch, M. M., & Loeding, D. Effects of position and type of question on learning from prose material: Interaction of treatments with individual differences. Journal of Educational Psychology, 1974, 66, 40-48.

diSibio, M. Memory for connected discourse: A constructivist view. Review of Educational Research, 1982, 52, 149-174.

Smith, E. E. Theories of semantic memory. In W. K. Estes (Ed.), Handbook of learning and cognitive processes (Vol. 5). Hillsdale, N.J.: Erlbaum, 1977.

Spiro, R. J. Remembering information from text: The "state of schema" approach. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, N.J.: Erlbaum, 1977.

Spiro, R. J. Constructive processes in prose comprehension and recall. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension. Hillsdale, N.J.: Erlbaum, 1980.(a)

Spiro, R. J. Prior knowledge and story processing: Integration, selection, and variation. Poetics, 1980, 9, 313-327.(b)

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

Swenson, I., & Kulhavy, R. W. Adjunct questions and the comprehension of prose by children. Journal of Educational Psychology, 1974, 66, 212-215.

Swets, J. A., & Feurzeig, W. Computer-aided instruction. Science, 1965, 150, 572-576.

Tannen, D. What's in a frame? Surface evidence for underlying expectations. In R. O. Freedle (Ed.), New directions in discourse processing. Norwood, N.J.: Ablex, 1979.

Taylor, S. E., & Crocker, J. Schematic bases of social information processing. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), Social cognition. The Ontario Symposium (Vol. 1). Hillsdale, N.J.: Erlbaum, 1981.

Taylor, W. L. "Cloze procedure": A new tool for measuring

readability. Journalism Quarterly, 1953, 30, 415-433.

TenBrink, T. D. Evaluation. A practical guide for teachers. New York: McGraw-Hill, 1974

Thomas, E. L., & Robinson, H. A. Improving reading in every class (Abridged 2nd ed.). Boston: Allyn & Bacon, 1977. (Orignially published, 1972.)

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

Thorndyke, P.W.,& Yekovich, F. R. A critique of schema-based theories of human story memory. Poetics, 1980, 9, 23-49.

Tierney, R. J., & Mosenthal, J. Discourse comprehension and production: Analyzing text structure and cohesion. Urbana-Champaign: University of Illinois. Technical Report No. 152, 1980.

den Uyl, M., & van Oostendorp, H. The use of scripts in text comprehension. Poetics, 1980, 9, 275-294.

van Dijk, T. A. Semantic macro-structures and knowledge frames in discourse comprehension. In M. A. Just, & P. A. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, N.J.: Erlbaum, 1977.

Walker, C. H., & Meyer, B. J. F. Integrating information from text: An evaluation of current theories. Review of Educational Research, 1980, 50, 421-437.

Warren, W. H., Nicholas, D. W., & Trabasso, T. Event chains and inferences in understanding narratives. In R. O. Freedle (Ed.), New directions in discourse processing. Norwood, N.J.: Ablex, 1979.

Watts, G. H., & Anderson, R. C. Effects of three types of inserted questions on learning from prose. Journal of Educational Psychology, 1971, 62, 387-394.

Winograd, T. Understanding natural language. Cognitive Psychology, 1972, 3, 1-191.

Winograd, T. A framework for understanding discourse. In M. A. Just, & P. A. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, N.J.: Erlbaum, 1977.

# APPENDIX

## SCRIPT OF INSTRUCTIONS GIVEN TO THE STUDENTS

We are students at McGill and we are trying to find out how children like yourself learn from science books they read in school. You can help us by doing some reading and telling us about it.

This is not a test and it doesn't have anything to do with your marks in school. You shouldn't worry about things like spelling and punctuation. We are trying to find out about how things you read in school are understood by sixth graders and about how to make these books better for you and your friends.

We are going to ask you to do some reading and some question answering today and some more next week at another session. Today we are going to give you two booklets to work on. First you have a couple of pages to read. Please read these pages carefully to yourself once. You don't have to hurry.

When you are through with the reading, you should go to the next page of the booklet and write in your own words what you remember about the passage. You should not look back at the reading when you are doing this.

After, you will find some questions. We want you to answer these questions one after the other without going back and forth.

OK? Anyways on each page of the booklet you are told exactly what to do. You'll find your way.

Once you are finished with the first booklet, you must answer the questions in the second booklet. These are the same questions as before but this time you can look back at the passage.

Ok? Do you have any questions?

Here are the booklets. Please do not start working until I tell you to do so. [DISTRIBUTE...]

First of all please put your name at the top of each booklet. Then, you can start working. If you are finished before your friends please wait. You can do some other work while waiting. Go ahead.

Debriefing

How did you find this? Comments? [WE ANSWER THEIR QUESTIONS...] OK. We thank you very much for your help.

One more thing. Since we will be coming back next week, please don't talk about the material you have read.