

32

MODELING THE PROCESSES HUMANS USE TO  
COMPREHEND WRITTEN SHORT STORIES

by

SHERNISE ALEXA SPEARMAN

B.S., Kansas State University, 1980

---

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Computer Science

Kansas State University  
Manhattan, Kansas

1983

Approved by:

  
Major Advisor

LD  
2668  
.R4  
1983  
.S65  
c.2

A11202 244596

TABLE OF CONTENTS

	<u>Page</u>
LIST OF FIGURES AND TABLES . . . . .	iii
ACKNOWLEDGEMENTS . . . . .	iv
INTRODUCTION . . . . .	1
SECTION 1 -- ANALYSIS OF THE PROBLEM . . . . .	3
SECTION 2 -- REPRESENTATION . . . . .	32
SECTION 3 -- IMPLEMENTATION . . . . .	39
SECTION 4 -- VALIDATION AND EVALUATION . . . . .	39
CONCLUSIONS . . . . .	42

**THIS BOOK  
CONTAINS  
NUMEROUS PAGES  
WITH THE ORIGINAL  
PRINTING BEING  
SKEWED  
DIFFERENTLY FROM  
THE TOP OF THE  
PAGE TO THE  
BOTTOM.**

**THIS IS AS RECEIVED  
FROM THE  
CUSTOMER.**

## LIST OF FIGURES AND TABLES

	<u>Page</u>
Figure 1. <u>Dick and Jane in the Park</u> . . . . .	3
Figure 2. Analysis of the Dick and Jane Story . . . . .	10
Figure 3. <u>Dick and Jane in the Park</u> showing the major propositions . . . . .	13
Figure 4. Story Tree of the Dick and Jane Story . . . . .	14
Figure 5. A partial Listing of Fillmore's (1971) Cases . . . . .	28
Figure 6. Four Partitions of Memory with Respect to Written Lanugage Comprehension, Sanford and Garrod (1981) . . . . .	30
 Table 1. Rules from Thorndyke (1977) Story Grammar . . . . .	 12



#### ACKNOWLEDGMENTS

I would like to thank my major advisor, Dr. Roger Hartley, and Professor Clifford Stark for their helpful comments and general enthusiasm for this area of AI research. I would like to thank Dr. Elizabeth Unger and Dr. Rodney Bates, my other committee members, for their help in writing this report.

Thanks also go to Dr. Richard Harris, who helped me understand the psychological view of language comprehension, and to Professor Mary Nichols for an extremely informative interview on the storytelling art form.

Last, but not least, I want to thank my family for all their love and constant support.

## INTRODUCTION--Language Comprehension in General

The adult human mind is designed to efficiently and accurately comprehend a person's native language. With rare exception, whether one understands spoken or written language this understanding process, through practice, becomes automatic, requiring very little conscious mental attention. One only becomes aware of some of the complexity of this process when the understanding process fails or falters--when comprehension becomes difficult. Ambiguous, anomolous, or paradoxical statements often cause such confusion in comprehension.

Linguists, psychologists, philosophers, computer scientists and educators have studied the human comprehension process for some time now, each deriving a different view of the process. All, however, have found that the true nature of a subconscious, mental process such as language comprehension is impossible to observe directly even with scientific experimental methods. One can only speculate as to what really goes on in the mind.

Science has turned its attention to modelling the comprehension process with machines (computer simulations) and to testing observable behaviors such as reading time, recognition time and recall accuracy with respect to a given written text from which inferences can be made about how well or to what degree of accuracy a given passage of text was comprehended. No one has yet completed the full picture on language comprehension. The Artificial Intelligence (hereafter AI) branch of computer science is developing several systems that partially model

the human comprehension process, each focusing on one type of story structure. Components used in these models vary greatly just as there is great variety in the theories backing the models. The general technique involves isolating some small facet of the total language comprehension process and attempting to construct a model of that portion that will rival human performance.

One such facet, written short story comprehension forms the basis for this report. Building a model for comprehending written short stories is no small task. Although the stories may seem trivial to the reader, it is only because the human comprehension system is already finely tuned and very efficient.

Discussion begins with analysis of the problem of "understanding" a short story including the views of human experts from different fields, each with a different perspective on the subject. Next, an introduction to the subject of story representation focusing on the processes involved in transforming written short story texts into some kind of form that somehow reflects the notion of "understanding" or "comprehending" the story. The report concludes with model implementation ideas and some discussion on the suggested criteria for the verification and evaluation of such a system.

## ANALYSIS OF THE PROBLEM

Consider the following short story:

Dick and Jane in the Park (figure 1)

Dick and Jane went to the park to play catch. They were throwing the ball to each other when along came a scruffy mutt named Spot. Spot saw the ball and wanted to play with it, so when Jane threw the ball in the air, he jumped up, caught the ball and ran away. Jane started crying. Dick ran after Spot. Finally he caught him by the tail; Spot howled. Dropping the ball, he scampered off. Dick picked up the ball and went back to play with Jane.

This little short story, perhaps at a second grade comprehension level, can easily be understood by an adult human. Yet what is the nature of this understanding? How can one test how well a story was comprehended?

One's comprehension of such stories is often tested by having to answer specific questions about the story. These questions generally fall into two basic categories:

(1) Questions about information or events explicitly contained in the text, such as--

- a. Who was playing catch in the park?
- b. What did Spot do?
- c. How did Dick catch Spot?

and (2) Questions about information or events not mentioned directly, but somehow inferred from the text, such as--

- d. Why was Jane crying?
- e. Are Dick and Jane friends?
- f. Why did Dick run after Spot?

These questions (2) require that the reader comprehended the story text itself and used his general knowledge about word definitions, relationships, events and rules of cause and effect to "read between the lines" of the story inferring knowledge that is implicitly associated with the story text. Deriving implicit knowledge is an essential part of the undersatnding. Writers will purposely omit information from the story that can be easily inferred by the reader.

Aside from the above questions, one is sometimes asked to do one or more of the following:

- (1) RECALL the text from memory;
  - (2) RECOGNIZE passages from the text;
  - (3) PARAPHRASE the text;
- or (4) give the GIST of the story.

Being able to recall the text, verbatim, requires that the entire text be stored in memory. Most people find this task to be difficult. Recall is best immediately after the reading and generally diminishes with the passage of time. Psychologists cite several probable reasons for this difficulty, chief of which is the fact that, by nature, one's memory is not designed to allow one to store exact copies of every incoming stimulus (in this case, written text). Short-term memory (hereafter STM) has a limited capacity which, if the text is too long, will overflow causing portions of the text to be "forgotten". Total recall is often facilitated by practicing good memorization techniques as one reads, such as rehearsing the phrases several times. This allows a more complete copy of the text to be encoded from STM to long-term memory (hereafter LTM). Chunking or grouping

the incoming text into meaningful grouping allows more information to be stored in STM. It is possible that the use of this phenomenon explains why stories that follow predetermined scripts or scenarios (a concept to be discussed in detail later in this report) are better comprehended or at least more easily understood. Psychologists also believe, and most researchers from other fields will agree, that comprehension alters the form of the text in such a way that total recall becomes difficult requiring extra effort to reconvert the text into its original form.

Recognizing passages from the text has been shown to be a less difficult task than the recall task. Psychologists feel that this is because the form of the recognition question offers the reader more ways to search memory for the answer. For example, after reading the Dick and Jane story, one might ask the reader:

g. Who ran after Spot?

This recall-type question gives the reader only one path to search for the answer: from the node\* representing Spot to the node representing Dick. Whereas asking the reader a recognition-type question like:

h. Did Dick run after Spot?

allows the reader to search either from Dick to Spot or from Spot to Dick to verify the fact that the story said "Dick ran after Spot."\*\*

Paraphrasing the text involves putting the text in one's own words. An example paraphrase of the Dick and Jane story might go:

---

\*The notion of a node will be explained in more detail in the discussion of propositional structures.

\*\*For more on the topic of Recall vs Recognition see pages 181-184 in Anderson (1980).

Dick and Jane were playing catch in the park when a dog named Spot jumped in and grabbed the ball Jane threw. As he ran away, Jane began to cry. Dick then ran and caught Spot by the tail causing Spot to drop the ball and run away in pain. Dick took the ball back to play catch with Jane again.

Research has noted a correlation between the interval of time between reading the text and paraphrasing it and the distortion of the paraphrase. Humans seem to distort the text by remembering events not actually in the text or even implied by the text. They also forget text-related information altogether (see Sanford and Garrod, 1980). The comprehension process activates memories of events, similar to those in the text, that one has personally experienced. Knowledge from these memories help one understand the text better, but such knowledge often gets incorporated into the text's representation in memory distorting it, making it more consistent with one's general knowledge or beliefs (see the discussion on scripts and scenarios later in this report).

Giving the gist of a story involves producing a capsule-summary or review of the entire story. It most often contains the major events in the story. The gist of the Dick and Jane story might go:

While Dick and Jane were playing catch in the park, Spot, the dog, came and took their ball, Dick got it back from Spot so that he and Jane could continue playing catch.

Unlike paraphrasing, the objective of gist is to summarize the main points of the story, detail is noticeably left out of the gist. Humans have a much easier time producing the gist of the story as opposed to

recalling details. This is especially true if the story is longer and/or if a great deal of time has elapsed between reading the text and being asked to give the gist or to recall the text.

Thus looking at the types of questions one uses to test or verify comprehension of a story, and looking at the resulting accuracy of the answers provides some insight into what kind of information is extracted from the text during the comprehension process and also how long and in what form the information is held in memory. It also shows the complexity of the comprehension process suggesting that previously stored knowledge and processes from LTM are needed to form that internal, mental representation of the story. Many of the questions asked about a story require implicit knowledge, knowledge not found in the written version of the story, yet somehow incorporated into the final internal version of the text.

There are several levels of analysis that come together to make up the total process of comprehension. The following is a list of some suggested divisions or levels of analysis:

- (1) GRAPHO-PHONEMIC--graphic patterns are analyzed into identifiable letters and special symbols (such as numbers and punctuation symbols) and then into phonemes;
- (2) SYLLABIC--phonemes are clustered into recognizable syllables;
- (3) LEXICAL--identifying words (joining syllables into words);
- (4) PHRASAL--syntactically grouping words into phrases;
- (5) SENTENTIAL--putting phrases together to form syntactically correct and semantically meaningful sentences;



(6) SUPERSENTENTIAL or CONCEPTUAL--grouping sentences into super-sentential groups of a more semantic nature, often relating sentences to events, plots, goals, etc. and other conceptual structures that are associated with stories.

Going from (1) to (6) represents a progression from low to high level analysis. These processes relate to each other, but not necessarily in a sequential manner. More often, in fact, they work in parallel. Low-level analysis will be facilitated by high-level analysis and vice versa. For example, several psychological experiments\* have shown that letters are better identified in the context of a word, and can sometimes even by-pass certain levels of analysis altogether as is the case when a word is comprehended without noticing that it is misspelled. When reading a story, the reader attempts to extract a meaningful representation of the text. This structure or mental representation of the story is formed through the interaction of the above-mentioned levels of processing.

Although at first this representation may include all the details of the story (something akin to a verbatim copy of the text), as time passes and processing continues, this representation is restructured. What remains is a structure that facilitates remembering the gist of the story, but discourages memory for details (with the possible exception of unusual or strange events). This more permanent representation of the text also contains new information, knowledge not explicitly stated in the story but merely implied by or inferred from the text. Sometimes these new facts distort the meaning of the actual story. The current

---

\*See the discussion of work started by Reicher (1969) and Wheeler (1970) in Sanford and Garrod, 1980.

ideas about the representation of the text will be postponed until later in this report. Now, let us concentrate on the high level analysis used to help form this mental text representation.

These processes, usually called Top-Down or Conceptually-driven processes, provide one with certain expectations about what is to happen next when a story is being read. They help fill in information left unclear or completely omitted by the lower-level processes.

For instance, most agree that stories have some underlying structure. Stories that conform to the structure are well-formed; stories that depart from the structure or that permute the structure are not necessarily ill-formed, but they are frequently harder to read taking more effort to comprehend. This suggests that one relies heavily on this structure to facilitate the process of story comprehension.

A well-formed story, such as the Dick and Jane story, has some basic characteristic components. For example, the story will have\*

(1) CHARACTER(s)--person(s) or thing(s) that are the central actor(s) in the events of the story,

(2) SETTING(s)--some kind of contextual; physical or mental background, a "backdrop" around which the story unfolds;

(3) CONFLICT(s)--action(s) that form the plot of the story. They are the basis upon which the story is built; without conflict it is difficult to understand much of a story. The story introduces the conflict(s) which then must be resolved through some action or scenes of actions which involve the characters of the story. The actor(s) that

---

\* Most of these components were suggested to me in an interview with Mary Nichols (Department of Speech, KSU), Fall 1982.

occur in an attempt to resolve the conflict are sometimes called goals; each goal is initiated (sometimes broken up into subgoals) and eventually either succeeds, fails or is abandoned in favor of another plan of action. This type of structure is called a goal-directed story. Conflicts are also tied to episodes. Usually there is one major conflict per episode; the episode concludes with the resolution of that conflict;

(4) PLOT--plot is closely related to (3) because the attempts at conflict resolution drive the plots of the story. A plot is made up of one or more episodes. It is also closely related to the notion of "gist"; describing the plot is a way of summarizing a story in a spatial-temporal order that frequently links the events in a causal-chain.

With these characteristics in mind, what follows is a rough analysis of the Dick and Jane story:

(figure 2)

title: Dick and Jane in the Park  
 characters: Dick, Jane, Spot  
 setting: the park

story (episode structure)\*

EPISODE I:

Dick and Jane play catch.

CONFLICT IA:

Spot wants the ball.

EPISODE II:

Spot plans to take the ball from Dick and Jane.

RESOLUTION II: (end CONFLICT I)

Spot takes the ball and runs.

CONFLICT IB:

Jane begins crying.

EPISODE III:

Dick runs after Spot.

CONFLICT III:

Dick tries to get ball from Spot.

RESOLUTION III: (CONFLICT IA ended)

Dick gets the ball, Spot runs away.

RESOLUTION I: (CONFLICT IB ended)

Dick goes back to play catch with Jane.

---

\*This format was adapted from a form used by W. Kintch in his analysis of stories from Boccaccio's Decameron in his paper "On comprehending stories" presented to Carnegie Symposium on Cognition, May 1976.

Moreover, stories that are well-formed seem to be abstractly divided into three subparts:

- (1) Beginning--where the majority of the main characters are introduced; setting is established, and at least one conflict introduced,
- (2) Middle--characters are developed and the conflict(s) progresses,
- (3) End--all major conflicts resolved, plot ends.

The beginning of the Dick and Jane story is the first sentence. The rest of the story is the middle except the last sentence, which constitutes the end.

As one learns to read stories, knowledge of these sorts of structures helps guide one through to a better understanding of the text. What happens if these structures are not followed? For one thing, comprehension suffers; that is, the comprehension process becomes more difficult or fails. Secondly, deviations from the above-mentioned story formats can cause the story to be remembered in a more distorted form with events occurring in the wrong places. Thirdly, if the departure from the norm is so novel, radical, or perhaps in some way shocking, the unusual events will remain focused in one's mind to the point that they actually become easier to recall than the ordinary events of the story. In order to increase the comprehension level, a writer who departs from the normal well-formed story structure leaves the reader clues or cues to help them restructure the story into its more normal sequence. This technique, for example, is commonly used in connection with story "flashbacks".

Researchers looking at the psychological aspects of story comprehension have differing opinions on how to incorporate such story superstructures into their psychological models of story comprehension. P. W. Thorndyke

(1977) incorporated some of the basic characteristic components of a well-formed story (discussed earlier in this report) into a series of grammar rules for simple stories. Those rules are listed in Table 1 below.

Table 1. Rules from Thorndyke (1977) Story Grammar.\*

Rule No.	Rule
1	Story $\rightarrow$ Setting + Theme + Plot + Resolution
2	Setting $\rightarrow$ Characters + Location + Time
3	Theme $\rightarrow$ (Events)* + Goal
4	Plot $\rightarrow$ Episode*
5	Episode $\rightarrow$ Subgoal + Attempts* + Outcome
6	Attempts $\rightarrow$ $\left\{ \begin{array}{l} \text{Event*} \\ \text{Episode} \end{array} \right.$
7	Outcome $\rightarrow$ $\left\{ \begin{array}{l} \text{Event*} \\ \text{State} \end{array} \right.$
8	Resolution $\rightarrow$ $\left\{ \begin{array}{l} \text{Event} \\ \text{State} \end{array} \right.$
9	$\left. \begin{array}{l} \text{Subgoal} \\ \text{Goal} \end{array} \right\} \rightarrow \text{Desired State}$
10	$\left. \begin{array}{l} \text{Characters} \\ \text{Location} \\ \text{Time} \end{array} \right\} \rightarrow \text{Statives}$
Legend	Note: A rule for Event was not given
$\rightarrow$	Can be rewritten as
( )	Content optional rather than necessary
*	Could be more than one of these
$\left\{ \text{or} \right\}$	Alternatives

\* Adopted from Sanford and Garrod's 1980 book, includes discussion of Thorndyke's work (pp. 74-76).

Thorndyke suggested that these rules be used or combined into a frame-like structure\* (after Marvin Minsky) called Read a simple story with slots for the various items listed in the rules (such as Time, Location, etc.), and each slot would have a specific set of selection rules associated with it to help extract or infer the proper information from the story to fill the slots.

This method called Read a simple story is based on the premise that the story must be broken down into its propositional form, a form proposed by another psychologist, Walter Kintsch.\*\* The story is then analyzed according to the rules in Table 1 to form a story tree. Figure 3 indicates the propositions used in the Dick and Jane story. Figure 4 shows the Dick and Jane story in tree form.

(figure 3) Dick and Jane in the Park  
with propositions indicated by numbers in parentheses

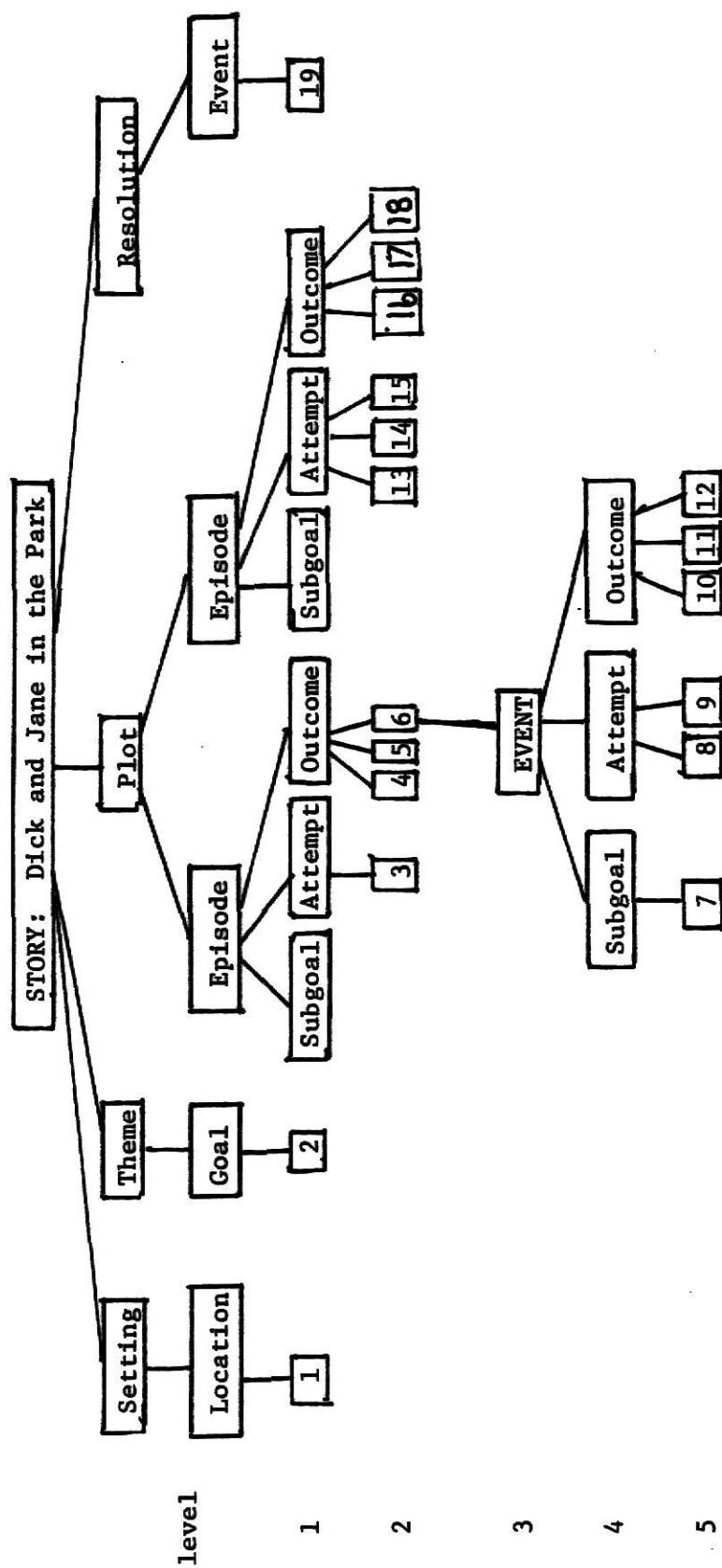
(1) Dick and Jane went to the park (2) to play catch. (3) They were throwing the ball to each other (4) when along came a scruffy mutt (5) named Spot. (6) Spot saw the ball (7) and wanted to play with it, (8) so when Jane threw the ball in the air (9), he jumped up, (10) caught the ball (11) and ran away. (12) Jane started crying. (13) Dick ran after Spot. (14) Finally, he caught him by the tail; (15) Spot howled. (16) Dropping the ball, (17) he scampered off. (18) Dick picked up the ball and (19) went back to play with Jane.

---

\* For a more complete discussion of frames with respect to story comprehension systems see Frame Conceptions and Text Understanding, Dieter Metzger (editor), Walter deGruyter (1980).

\*\* Propositions will be discussed later in this report.

(figure 4) Story tree of the Dick and Jane story, propositions indicated by numbered boxes.



The theory is that going from left to right at level 1 in the tree produces the "gist" of the story. Psychological testing showed that the lower the proposition was in the tree structure (i.e., the higher the level number), the more difficult it was to recall the proposition.

David E. Rumelhart suggested in his 1975 work that there is both a syntactic and a semantic interpretation (or representation) for a story. These representations were formed according to story grammar rules of the form: syntactic rule -- semantic rule (see pages 219-225; Bobrow (1975)). Rumelhart indicated his approach relates closely to the approaches of Roger C. Schank (as stated on pages 237-272 in Bobrow (1975) and Robert P. Abelson (as stated on pages 273-310 in Bobrow (1975)) only with less emphasis on Bottom-Up parsing (data-driven) and more emphasis on Top-Down (conceptual-driven) parsing of stories.

All agree, however, that well-formed stories have some kind of conceptual structure which guides one's expectations about what's to come in a story and helps one understand the text. Psychological research indicates that a story is more easily comprehended if the theme or major conflict around which the story is built is mentioned very early (the first few sentences) in the story. In the Dick and Jane story, the title and first sentence set the context and mood for the story introducing two of the three characters, the setting, and the major theme. If the story's theme was placed at the end of the story or left out entirely, comprehension suffered measurably.\*

One's understanding is also based on how well one understands the events of the story. For example, if one knew nothing about "playing

---

\* See a summary of Thorndyke's research on pages 78-80, Sanford and Garrod, (1980).



catch" or "dogs playing with a ball", then Dick and Jane in the Park would not be very deeply comprehended. Events, however, like playing catch for most bring to mind some stereotyped situation perhaps from one's childhood when one played catch with a friend or a parent. If asked to do so, one could come up with a list of characteristics that make up the idea of "playing catch", such as:

- (1) Catch is a game;
  - (2) It is played by two or more people with one person at any given time designated as "thrower" of the ball and the other(s) being "catchers" (or at least potential catchers) looking to catch the ball; or else if more than two are playing, some will be "watchers" (neither catching nor throwing, but still part of the game). Notice that if an animal participates in the game the name is usually changed to Fetch;
  - (3) Props used--usually a ball (optionally a glove is used to catch the ball);
  - (4) Major actions--throw, toss;
  - (5) Objectives or Entry conditions to the game--seeking pleasure, enjoyment, or exercise;
  - (6) Location--a clear area;
  - (7) Results or Terminating Conditions for the game--darkness (can't see to catch the ball), fatigue of participants, higher priority event needs attending to, or boredom (too many repetitions of the major action);
  - (8) Point of View . . .
  - (9) Time Constraints . . .
- etc.

These values which characterize a situation such a "playing catch" come together to form a script (Schank and Abelson, 1977). A script is a representation of the characteristics of a predictable situation or sequence of events. They are sometimes called schemata, or scenarios. Scripts become predictable because the same sequence of actions has been repeated several times until one comes to expect them to occur together. A script's purpose is to detail these events so that if one occurs in a story the reader will recognize that the events being read fit a script and then the reader can retrieve the script from LTM to facilitate comprehension as discussed before. The level of detail varies from theory to theory some using sentences, some using propositions and still others using primitive acts. Schank developed a series of such primitive acts which came to be known as his Conceptual Dependency Theory (see Schank, 1975), a theory of the representation of meaning of sentences.

Schank's CD theory did for the semantic representation of sentences what Noam Chomsky's (1962) transformational grammar theory did for the syntactic representation of a sentence. The basic axiom of Schank's theory is that sentences with identical meaning should have the same Conceptual Dependency (hereafter CD) representation. It follows, then, that any information that is implicit in a sentence must be made explicit in the CD representation (called a conceptualization). There are two kinds of conceptualizations:

- (1) ACTIVE, with the form --

Actor Action Object Direction (Instrument),\*

- (2) STATIVE, with the form --

Object (is in) State (with Value).\*

---

\*Material enclosed in parentheses is optional.

The CD theory in Schank and Abelson (1975) contains 11 primitive acts or actions:\*

- (1) ATRANS -- transfer of an abstract relationship such as control, possession, or ownership;
  - (2) PTRANS -- transfer of an object;
  - (3) PROPEL -- application of physical force to an object (a movement (PTRANS) need not take place;
  - (4) MOVE -- movement of a body part of an animal by that animal;
  - (5) GRASP -- grasping of an object by an actor;
  - (6) INGEST -- taking in an object by an animal to the inside of the animal;
  - (7) EXPEL -- expulsion of an object from the body of an animal into the world. Note that INGEST and EXPEL often work as inverse acts;
  - (8) MTRANS -- transfer of mental information between animals or within an animal. Memory is again partitioned into what Schank calls the conceptual processor (or CP), which is analagous to the psychological STM, and the long term memory (LTM). The CP or STM is where thoughts are generated; the LTM is where general knowledge is stored;
  - (9) MBUILD -- construction by an animal of new information from old information;
  - (10) SPEAK -- actions used to produce sounds;
  - (11) ATTEND -- action of focusing a sense organ toward a stimulus.
- These primitive acts are not verb category names; they are primitive action labels that will work for any language, describing the underlying

---

\*There is still some discussion as to the actual appropriate number of primitive acts.

actions that verbs, states and nouns represent. Using these primitive acts to symbolize the meaning of a sentence reduces the amount of inferencing since the inference rules are written once for each act and not once for every verb.

When a conceptualization involves an attribute-value statement a scale is used. Scales help indicate a change in state. The scale used in Schank and Abelson (1975) is a number continuum from -10 to 10. The numbers are, however, arbitrary and do not represent any actual quantitative relationship. For any given state, there are various words that have an associated number on the scale continuum. For example, the mental state "elated" might rate a 9 and "heart-broken" might rate -10.

According to the theory, sentences in a story are translated into their respective CD representations as they are read. These representations are linked together in a causal network\* according to a set of causal syntax rules. The resulting network can then be used to answer questions about the causes of events in the story. The five main rules are:

- (1) Actions can result in state changes;
- (2) States can enable actions;
- (3) States can disable actions;
- (4) States (or acts) can initiate mental states;
- (5) Mental states can be reasons for actions.

Their computer model of story comprehension, called SAM (1975), bases its comprehension on scripts. It builds a copy of both the causal network

---

\*For more on causal relationships see Chapter 2, Schank and Abelson (1975).

and a copy of the instantiated script used to understand the text.

Although the CD theory just discussed is an example of low-level (data-driven) processing, it was essential to talk about it because SAM accepts incoming text sentences in their CD representations. The scripts themselves are even written in terms of conceptual dependencies. Pattern matching or deciding whether or not a given sentence fits the chosen script for a story is made more efficient this way; thus, processing time is saved. This method, however, is an idiosyncrasy of this computer model and is not necessarily what actually occurs in the human mind.

In general scripts are listed in terms of the actions that occur. They also include information about the characters (their roles) and reasons for the various events in the script. Scripts indicate the order in which the events of a situation are carried out making it easy once in a script situation to make predictions about what is going to happen next. This is called Expectation-driven comprehension. The reader makes predictions about what's to occur next in a story. Those predictions are added to the context in which the text that follows will be understood. For example, knowing that dogs like to play with a ball, one can predict that Spot will try to take the ball away from Dick and Jane. This helps one interpret the passage:

" . . . , so when Jane threw the ball in the air, he [Spot] jumped up, caught the ball and ran away . . . ."

as being Spot's plan to get the ball.

Scripts also help to fill in the gaps between events actually in the story and events that were not explicitly mentioned in the story.

Consequently, if when the reader is asked,

i. What kind of animal was Spot?,

responded that "Spot was a dog.", it is because the inferences that "a mutt is a type of dog." and "a dog is an animal." are made by the reader while reading the story since the word 'dog' was never explicitly mentioned in the text.

It is important to note that the discussion, so far, about scripts was directed more toward the AI interpretation of the concept. Other disciplines are studying human story comprehension in an effort to confirm that such a knowledge structure does indeed exist in the mind. AI, by its very nature, is more interested in constructing a concrete version of the script knowledge structure that can be incorporated into a working comprehension computer model. Other disciplines, psychology in particular, use the notion of a script to design experiments to test human comprehension empirically which will imply the use of scripts.

Hence there are many differing opinions as to the length, detail, and characteristic features of a script. Some say that scripts should describe the events and characters involved in great detail, others contend that such detail is not necessary inside the script because inferencing, logic and deduction routines plus the wealth of general knowledge all stored in LTM can be called up when needed to fill in the necessary details.

Yet scripts are not the only high level structure used to improve comprehension of the story text. The Dick and Jane story provides a perfect example. In the middle of the script, "Playing Catch", a new

conflict is introduced that is not part of the typical catch scenario: a dog named Spot wants the ball. At this point the focus of the story shifts from Dick and Jane to Spot and his plan to get the ball. Spot's actions, however, do not follow a script; that is, one cannot exactly predict what action Spot will take to get the ball. He could bite the thrower's leg to make him drop the ball, then grab the ball and run. He could sit quietly whimpering until one of the characters felt sorry for him and gave him the ball. The choices are endless, yet each plan has a structure not based on predefined stereotyped events, but based on goals. In this case the goal is to get the ball, a goal that can be accomplished by a plan. Spot's plan of action becomes:

- (1) Wait until the ball is in the air;
- (2) Time the jump to catch the ball in the air before it reaches the catcher;
- (3) Catch the ball;
- (4) Run away so that they don't take the ball away.

When such plans of action or plan sequences become so stereotyped and predictable through constant repetition they are transformed into scripts. To reiterate, a plan is a sequence of actions or methods directing a character toward some ultimate goal. They take over when no script is available to help comprehend a situation. Just as Spot's plan centered around the goal of getting the ball, Dick's subsequent retrieval plan is centered around the goal of retrieving the ball from Spot.

Planning can basically be broken down into a five step process:

- (1) Ascertain the goal;

- (2) Formulate the plan or method of obtaining the goal;
- (3) Initiate the plan;
- (4) Evaluate the results

IF goal achieved, THEN "success", go to 5

ELSE (a) Reaccess plan and/or goal

(b) Ask-can plan and/or goal be restructured?

IF yes THEN do so and REPEAT (1) with the new  
plan/goal

ELSE "failure", goal not achieved, go to (5)

- (5) Terminate plan.

Plans can be used to define verbs and common actions in general. Named plans are plans that, like scripts, have become the predictable sequence of actions yet not to the high degree that scripts are predictable. Nonetheless the plan sequence is given a name. The sequence represents the usual path the attainment of the ultimate goal.

So at a deeper, more fundamental level, text comprehension involves identifying and keeping track of goals to determine whether it is obtained (success), restructured, not achieved (failure) or abandoned. The goals of a character in a story effect his actions. The reader must be aware that the state or outcome of a goal alters the meaning of the story and also effect how scripts and plans are used to form the reader's expectations. A simple but illustrative example is the effect of Dick failing to catch Spot (a subgoal of his ultimate goal: to retrieve the ball). Indeed humans consciously test themselves on this ability to monitor the effects of goals every time someone starts a question with "what if . . .?"



Schank, Abelson and their students at Yale University developed a computer program, PAM (Plan Applier Mechanism), that would handle stories that use plans like those mentioned above.

Another high level knowledge structure is the theme. Themes are usually triggered by key words in a story. They help one set up expectations, for example, about a character's behavior. One would expect a lawyer (theme) to act different in a restaurant than a bum (theme). This type of theme is called a role theme because it characterizes the role a character plays in a story. Other themes, such as relationship themes designate the prescribed interaction between characters as in the MOTHER/DAUGHTER theme. Themes about a character's ambitions, like DESIRES TO BE RICH, will help the reader understand the goals that character sets for himself. With the expectations and belief triggered by these high level knowledge structures help the reader "get into" the characters and their actions, thus promoting a deeper understanding of the story text.

In summary, one's understanding of a story can be guided by:

- (1) Beliefs,
- (2) Expectations,
- (3) Themes,
- (4) Goals,
- (5) Plans,
- (6) Scripts, and
- (7) Story grammars.

These kinds of general world knowledge along with other factual information reside in LTM until they are accessed to be used to process the story. One can talk about levels of comprehension according to the use of the above-mentioned structures.

Beliefs and Themes are used mainly to check the characters and their actions to make sure they are consistent with the belief system of the reader and the themes associated with the characters. Scripts and story grammars give the reader a feel for the sequence of events in the story, the overall structure of the story. They also help the reader form expectations by being able to look at the script or grammar and predict what will happen next. But if the reader is to most fully comprehend a story he must understand, to be able to explain or justify, the plans and goals which direct the actions of the characters. This means that the answer to

j. Why did Jane cry?

requires a deeper understanding of the Dick and Jane story than does the answer to

k. Did Jane cry?

Answering (j) requires that one know the reason or goal behind Jane's crying, whereas (k) simply tests one's ability to recall part of the story.

Deeper understanding means more processing time. Answering (j) would take longer than answering (k). Studies have shown that such deep processing leads to better comprehension and longer retention of the story in memory. In general, the shallower the depth of processing, the more superficial the comprehension of the text.

The knowledge structures just discussed constitute the Top-Down, Conceptually-driven aspect of story comprehension. They frequently

interact with each other and often work in parallel to the lower level processing. It was earlier mentioned that translating sentences into CD form was one such low level process.

The reader starts the whole comprehension process by reading the text off the page. This approach to comprehension, called Data-driven or Bottom-Up processing,\* starts at the sensory input level of reading and recognizing the printed letters. The letters are then pieced together to form phonemes, then syllables, then words, phrases and eventually sentences. Emphasis is not only placed on the syntactic correctness of the written material but, by interacting with the more conceptual levels of processing, the semantic meaning is also formulate. Ultimately the memory for the exact syntactic form of the text is forgotten and only its semantic meaning is retained in memory for any length of time. Therefore the goal of comprehension, a mental representation of the story, reflects how well the meaning of the story was comprehended.

Opinions vary on the final form of this mental representation. Schank and Abelson's opinion was discussed earlier with respect to script-based understanding. Looking at the cognitive psychologist's interpretation, most begin by saying the sentences are broken up into propositions. As proposed by Walter Kintsch (1974), a proposition consists of a predicator and  $n$  arguments ( $n \geq 1$ ). These predicators and their arguments consist of word concepts occurring in or derivable from the text. Valid combinations of predicators and arguments are determined by the lexicon entries for the various word concepts which contain

---

\*Low level processes will not be elaborated on in this report.

the meaning of and the uses for all the word concepts known to the reader. A "word concept" is a more abstract term than "word"; it can be realized in the surface structure of a sentence as a word or as a phrase. Kintsch (1974) felt that word concepts get a fair amount of their meaning from the context in which they are used. Therefore, in the lexicon, it is not necessary to specify the meaning in great detail as long as mechanisms are provided to elaborate on this meaning in a particular context. The lexicon is part of semantic memory which contains one's general world knowledge. A particular word concept, for example CHAIR, is stored in the lexicon and called the CHAIR type. This word concept can appear as a token (similar to the concept of instantiation) in other word concept types. For instance, the CHAIR token may appear in the FURNITURE type. This token is "understood" by referring back to its type.

Texts are represented by word concept tokens. Understanding a text first involves understanding the tokens by referring back to the lexicon to find their types and then interpreting the interrelationships between tokens. Kintsch prefers also to leave the representation of the text in propositional form instead of further decomposing the propositions\* into semantical primitive acts for the following reasons:

(1) It is hard to see where decomposing should stop. No one has provided a list of primitive acts from which every concept can be reconstructed. Often such a list will lead to decompositions that only show one aspect of the word concept's meaning in a given context;

---

\*Psychological evidence points toward propositions being multivalued (i.e., perhaps based on a system of certainty factors); however, some comprehension theories and models give propositions simple true/false values since they were derived from first-order predicate calculus and propositional logic.

(2) It seems enigmatic that such decompositions are necessary for representing word concepts in memory and for comprehension since language has evolved to the point where complex word concepts are used;

(3) Chapter Eleven in Representation of Meaning in Knowledge and Memory (Kintsch, 1974) presents data reported which does not support the notion of decomposition.

If the predicator represents a verb action or adjective description, Kintsch uses Fillmore's (1971) case grammar cases (partially listed below in figure 5) to identify the relationship of the argument to the predicator.

(figure 5) A partial listing of Fillmore's (1971) cases.

agent (A)	-- the instigator of an action
experiencer (E)	-- the experiencer of a psychological event
instrument (I)	-- the psychological stimulus of an experience or action elicitor
source (S)	-- the source of an action in space, time, or a transfer event
goal (G)	-- the goal or result of an action
object (O)	-- the object of an action which undergoes change or movement.

The sentence

1. The man broke the window with a stone.

would be written:

m. (BREAK, A: MAN, I: STONE, O: WINDOW).

The sentence

n. A dog is a pet.

involves a nominal predicate, which are treated differently. Its representation would be

(figure 6) Four partitions of memory with respect to written language comprehension, Sanford and Garrod (1981).

	Dynamic	Static
Text-based	Explicit focus	Long-term text memory
Knowledge-based	Implicit focus	Long-term semantic memory

where,

(1) Explicit focus contains the events and entities explicitly mentioned in the text. This area of memory has a limited capacity (further discussion in Chapter seven of their textbook);

(2) Implicit focus contains the current subset of general knowledge along with any currently active scenarios (a scenario is very similar to a script) which is helping interpret the text;

(3) Long-term text memory contains the LTM representation of the text (a subset of episodic\* memory);

(4) Long-term semantic memory contains the general knowledge base (often just called semantic memory).

---

\*Episodic memory is memory organized around one's own personal experiences and including spatio-temporal references.

o. (PET, DOG).

Kintsch's complete theory includes mechanisms for inferencing and deduction needed to extract implicit as well as explicit knowledge from the text when creating its propositional form.

Kintsch chose not to use an associative network representation\* to show the relationships between the text's propositions preferring instead to apply a system of rules to the text base to generate these relationships when and if they became necessary in the comprehension process.\*\*

Sanford and Garrod (1981) developed a "processing framework" based on experimental results of reading comprehension tests which they hoped shows a clear relationship between knowledge access and comprehension in the human reader.

When looking at the process of text comprehension, they basically considered memory as being divided into at least four partitions shown in figure 6 below.

They discussed how knowledge is retrieved from and constructed in memory in terms of two primitive procedural functions:

(1) CONSTRUCT: which incorporates new entities from the text or from inferences made from filling slots in the scenario into memory, in particular into explicit focus; and

(2) RETRIEVE: which helps resolve references in the text on the basis of

(a) domain--which focus should be searched, or if such search fails, using a search of LTM;

---

\*See Anderson (1980) for a description of a propositional network.

\*\*For a more detailed explanation of his argument for this rule-based approach to representation of the text base in episodic memory see Kintsch (1974).

- (b) partial description--some noun phrase, pronoun or decomposition of a token to aid the search;
- (c) return:token (used here in the same sense as Kintsch (1974)

identity (if explicit focus is the domain), or  
slot identity (if implicit focus is the domain).

The return value specifies the entity or more accurately the pointer to the entity to be returned from the retrieve function.

Sanford and Garrod elaborate on these two operations in Chapter Eight. The resulting LTM text representation is a combination of all that was used in explicit and implicit focus during the comprehension process. They favored a script or scenario-based comprehension system although the emphasis of their book was on the psychological backing for such a system and not on the particular about what a scenario should contain or how it should be represented in memory. Like Schank and Abelson (1975), their model of comprehension was based on the readers ability to retrieve scripts, from cues in text to aid the transforming the text into its mental representation.

In general, no matter what the representation of the story looks like it should reflect the use of one's general knowledge in the interpretation of the text and should account for such phenomena as:

- (1) the increased inability to recall the details of the story as time progresses;
- (2) the ability to paraphrase a story;
- (3) the ability to give the gist of a story with relative ease (as compared to recalling details);



(4) the ability to recall material not explicitly mentioned in the story but implicitly related to the text if the model is to begin to rival the actual human comprehension process.

Although theories exist as to how these cognitive processes work, they are at best theories that were derived from inferences made from human comprehension test data. Thus simulating these processes requires a great deal of speculation and guesswork to fill the gaps in the theories caused by our limited knowledge. A computer model will only be able to model a small subset of the possibilities that a human is capable of handling. That set will restrict the text studied to simple, well-formed short stories. The notion of a well-formed story has been discussed earlier in this report. "Short" means stories with 50 propositions or less. "Simple" means all the knowledge in the story and the general knowledge needed to interpret the story must be representable in some form in the computer (presumably in either propositional or conceptual dependency form).

#### THE REPRESENTATION

During the course of the understanding process, the text must go through at least the first three of these four states:

(1) the ORIGINAL TEXT (as it's read in; syntactic structure preserved);

(2) the PARSED VERSION of the text (in cononical form, either propositional or CD conceptualizations (semantic meaning preserved). This version includes all inferred information gained when the text is processed and is dynamically changing in STM as each sentence is typed in;

(3) the RECONSTRUCTED (Final parsed) VERSION -- As time progresses and because episodic memory space is limited, version (3) of the text will eventually lose its spatio-temporal relations\* that characterize all episodic memories and will be transferred and perhaps restructured into long term semantic memory where more "general" knowledge is stored.

Psycholinguists in their attempt to construct a theoretical system that would accommodate these various text versions have agreed upon at least four basic components needed in the system:

(1) INTERFACE PARSER -- to convert the input word strings into a memory representation. This parser should be "intelligent" enough to handle common input errors, such as spelling mistakes or pressing the wrong key as the text is entered from a console;

(2) REPRESENTATIONAL STRUCTURE -- to embody one's general knowledge about language and the world which will be needed to interpret the text. This structure or memory should be partitionable so that at any given point in time during the processing of the text only a small portion of the structure could be accessed by the interpretation routines. This makes memory searches for information much more efficient and also simulates the notion of both focus and point of view in the story. The partitions must be dynamically changeable as more text is read and interpreted;

(3) CONTROL SYSTEMS -- to monitor and guide the memory searches, interfacing process, etc. These systems or procedures should incorporate

---

\*Any references, for example, to when the text was read and under what environment conditions.

the concept of a belief system so that inferences and text sentences can be checked for inconsistencies. These procedures help formulate the expectations that will facilitate comprehension.

(4) OUTPUT SYNTHESIZER -- for converting the mental information into spoken, written, or physical responses. In particular this component should act as an answer generator for questions proposed to the system, since questions are a handy way of verifying how well the text was understood.

The human mind contains efficient, fairly rapid, accurate retrieval and search procedures for LTM (both semantic and episodic). The inferencing procedures are especially efficient because they are used frequently. Much of the text, by the very nature of natural language, is ambiguous, incomplete or redundant. The human comprehension process relies heavily on the inference mechanisms to eliminate the redundancies or disambiguate the input strings. The internal text string representations are designed to facilitate inferencing.

A computer comprehension system must try to simulate these properties. Ideally for the system to be complete, it must also be able to "learn" new information such as new words, phrases or concepts by either recombining portions of current general knowledge (known in internal memory) or by accepting additional input (knowledge) typed into the system by the user.

From an AI point of view, a computer system would start out by taking the text in its English form and parse each sentence, one at a time, into its "primitive" form (either using propositions or CD conceptualizations). This form makes it easier for the system to do pattern matching and inferencing because:

(1) it eliminates all the peculiarities of natural language syntax structure; and

(2) all the information in memory, text and general knowledge, will then be in the same form.

Since this translation occurs one sentence at a time, it is not possible to solve any ambiguous references at that time; but these references should be tagged so that later, when more of the meaning of the text is known, these references can be resolved. Inferencing, too, must be postponed due to the lack of information at the sentence level.

Once a sentence is in its primitive form, a procedure would scan it to see if any of the conceptual knowledge structures (scripts, plans, goals, or themes) are cued in the sentence. The idea is to find such a cue so that the structure can be used to predict or explain the actions of the characters.

One might first scan for a cue to a script because a script will give the system the most information as far as predictability of actions. To facilitate this process, an index for each of the four knowledge structures should be kept that lists all the structures of a given type that are known by the system and any rules needed to locate the cues in the primitive form. In this case the index would be for all known scripts. Generally scripts are cued by the presence of the name of the script.

If a script is not found another procedure will be needed to search for a goal. Recognizing a goal is more complicated because the cue probably won't be a single word (except perhaps in a single-word theme like LOVE). Themes about the aspirations of a character have associated

probable goals, so finding a theme will be a start. Another way to find a goal is to examine the actions for their causal structure.

Once a script is found to guide the interpretation of the text, several things can happen. The first will be to instantiate as much of the script as possible by using the sentence (in primitive form). The sentence won't always match the first action of the first event of the script. If this is the case, then all the actions leading up to the one matching input sentence are assumed to have already occurred. This is because the events of a script are structured in their causal order. These actions must be instantiated and added to the text's internal representation (the instantiated script). Also at this point, any references or inferences that can be made should also be added to the text's representation. Finally, a set of predictions about the actions and/or events to come in the story should be generated in such a way that they will aid in interpretations of the next sentence. This process should be repeated until one of two things occur:

- (1) There are no more input sentences (i.e., the entire text has been processed and instantiated by one known script); or
- (2) A sentence has occurred that cannot be interpreted by the currently active script.

In the former case, processing is complete and various methods of testing the comprehension accuracy can begin. In the latter case, it is possible that:

- (a) a cue for another script was found,
- (b) a cue for a goal (i.e., plan) was found, or
- (c) the sentence cannot be interpreted by the current system.

In case 2a, the current script must be replaced at least temporarily by the new script (this process is analogous to the phenomenon called "shifting focus in a story"). The now deactivated script should be "shelved"\* in some storage location with a marker to show at what point the departure from the script occurred. If focus ever shifts back to script, then it will start its interpretation from that point in the script. Processing would continue by linking the instantiation of the old script to the new one, and then repeating the analysis steps as before to interpret the input sentence; this time using the new script.

In case 2b, the current script would again be deactivated, marked, and shelved, but this time control would be given to the planning routine.

If the system cannot fit the action of the character(s) into a stereotypical situation (using a script), then the system must focus on trying to explain the intentions of the character(s). This is done by trying to locate a goal which would cause such an action. The plan procedure will start by determining if a goal is indicated by the input sentence. Once this is done, memory must be searched to determine the best method or plan to achieve the goal. Next, the input sentence should be scanned to see if there is any evidence that the method selected is indeed being used by the character to achieve the goal. If this is true then the action has to be explained, and the system will continue instantiating the plan by checking to see if each of the subgoals of the plan are realized in the subsequent input sentences. This process

---

\*If the indexed list of all scripts was implemented as a stack then shelving would occur at the top of the stack thus giving the newly activated script first priority for choices of scripts to active again if another shift in focus occurs.

continues until the ultimate goal is realized or a new goal is indicated. If this happens, then the processing will shift to the new goal but still save the old goal by indicating at which subgoal the processing shifted and putting it back into the plan index. Processing would continue with a new plan and goal.

If the system ever fails to interpret a sentence (case 2c) one alternative would be to put the sentence aside and try to interpret the next sentence. Then if the new sentence can be interpreted the system could go back and try to interpret the other sentence in context of the new sentence. If this did not work then perhaps the process could continue as long as the sentences being read in could be interpreted. This process could theoretically go on until the whole text was understood except that uninterpreted sentence. A "failsafe" procedure like this could keep the system from coming to a crashing halt if a sentence could not be interpreted. In fact, every uninterpreted sentence could be stored and processed as such. When the entire text was processed, the user could then tell which sentences were not understood.

By combining the methods of script instantiation with planning, this hybrid system will greatly increase the number of stories that one system could interpret. The end result is a text representation (in terms of propositions or conceptualizations) in memory that reflects the use of the conceptual knowledge structures (especially scripts, plans, goals, and themes) in its structure.

In order to fulfill the criteria of a good comprehension system, it should be able to now paraphrase the story or to give the gist of the story. Procedures to do these things would have to traverse the final

text representation to create the necessary statements. Traversal procedures could also be written to answer questions about the setting, the theme, the main episodes or plot and the resolution of the story thus incorporating some of Thorndyke's (1977) ideas about the structure of the simple short stories.

## IMPLEMENTATION

Since the system just discussed is just a speculation, no implementation details will be given here.

A system that understands script-based stories, SAM (Script Applier Mechanism), already exists as well as PAM (Plan Applier Mechanism), which understands plan or goal based stories. There is also a system that translates English sentences into their equivalent conceptualizations. All of these systems are described in detail in INSIDE COMPUTER UNDERSTANDING: Five Programs Plus Miniatures, edited by R. C. Schank and Christopher K. Riesbeck.

At this time, research is underway to experiment with the interactions between SAM and PAM to perhaps produce a hybrid, more complete comprehension system as suggested above.

## VALIDATION AND EVALUATION

Testing or validation and evaluation of the theoretical system would consist of having a human comprehension expert either type in questions to the system about the text (if the text is processed as a whole before questions), or observe the kinds of questions the system generates when trying to



clarify the text (if the text is processed interactively, with aid from the expert).

In the interactive environment, the system could generate questions if it ever encountered a sentence it couldn't properly interpret. The expert would then note how relevant the question was and whether its answer will alleviate the problem. If it will, the expert should type in the answer to the question, the system then should use the answer to interpret the sentence, modify the text's representation, and continue processing. Questions could range from not knowing a word to being unable to resolve a reference to having the story somehow contradict the system's built-in system of beliefs (a collection of rules about how the world should "logically act"). An example contradiction would be if the following sentence was added as the very end of Dick and Jane in the Park:

"Dick gave the ball to Jane and she ate it."

The act of "eating the ball" would contradict the system's belief rules. It would state in effect that humans eat only food. The point is that not only must the problem be located by the comprehension system; but it also must be able to locate information needed to formulate the proper question to resolve the problem (all these require a great deal of sophisticated systems design). The expert, while testing the system, must monitor the performance level of both of these tasks.

Although the interactive question system would be a much more complicated and sophisticated system, the sequential mode (waiting until the entire text was processed) has some heavily weighted advantages.

First, it is a simpler system because it needs no interactive question/answer routines or components. Second, in order to evaluate the system's performance in comparison to that of a human, the methods of testing need to at least be very similar.

In the human experiments with reading comprehension, psychologists used the scientific measure called reaction time (or RT) for many of the experiments. It involves measuring, for instance, the time it takes the human to comprehend ("process") the given text (by reading the entire text, with no interactive questions allowed). Thus the interactive mode of processing would not allow the expert to use the wealth of already established ways of testing comprehension which use the RT measure.

Using the sequential method of processing, the expert could measure comprehension time as well as ask the system questions similar to those discussed early in this report with respect to the comprehension of the Dick and Jane story measuring the time it took the system to respond to each question (the system's "reaction time"). The accuracy of the response is also an important test.

The text representation should be verified both for its causal chain of events and for its spatio-temporal chain of events. In the testing stages of implementation, each primitive form (proposition or CD conceptualization) should be checked for accuracy also.

Other psychologically based questions that should be incorporated into the tests include asking the system to:

- (1) recall parts of the story;\*

---

\*Accuracy in these tests is generally based on the number and accuracy of propositions recalled. Exact syntactic form is not required.

- (2) recognize if given statement(s) were part of the story's text;
- (3) paraphrase; or
- (4) give the gist of the story.

Attempting to simulate the experiments psychologists use to test humans for story comprehension will help assure that the results are objective and accurate as well as making the results easier to compare with the actual human performance results.

Yet it must be remembered when comparing the results of humans to the results of the comprehension system, that there will have been several assumptions and liberties taken with the established theories of human comprehension in order to develop a complete, working computer comprehension system. Consequently, the methods used by the system to understand the story will undoubtedly be different than those used by humans (which at this point, one can really only speculate about). However, by graphing the performance results of the computer system in the same manner as done in various psychological tests with human performance data (reaction time experiments, in particular\*), perhaps some interest correlations between the relative shape of corresponding graphs for any given test might be found.

## CONCLUSIONS

Using the computer to actually model the complete process of human story comprehension is, at this point in time, still just a dream.

Some of the techniques used to simulate the components of such a system are still in their infancy; some have yet to be designed. The

---

\*See Sanford and Garrod, 1980, and Anderson for explanations, references to and experimental results from several psychological text comprehension experiments.

basic approach taken now by AI researchers is to select a specific kind of story, like one based on scripts or one based on plans and goals, and then to create a model that will theoretically process all stories of that type. For experimental purposes most of the stories used to test the systems are short (less than 10 sentences) and don't have complicated plots. One of the reasons the latter restriction is placed on the test stories is that all the general knowledge needed to understand the story, like the scripts, plans, goals, themes, etc., and inference rules, must be designed and built into the system by manually by humans. The systems do not contain learning procedures to generate new general knowledge internally. So to add a new theme, say JEALOUS, everything in the system affected by the theme must be manually redesigned by a human and then placed back into the system. The problems with complicated plots are:

- (1) the amount of work involved to design the knowledge structures, procedures and rules to handle them;

- (2) some of the knowledge needed may not be clearly defined or may be impossible (given present techniques) to design for computer use. Comprehension systems often fail to comprehend part of a story because they do not have all the necessary knowledge (especially the conceptual knowledge structures) available to them. Indeed, for any story comprehension system to be highly successful, it needs all these structures available; so that, if the system fails to comprehend a sentence using one structure, it can try to use another. This process would intuitively seem to mimic human's attempts to comprehend a story.

The study of the human language comprehension process, especially in the fields of cognitive psychology and psycholinguistics, is only now beginning to make some headway toward formulating a more complete model of how our subconscious mind works to comprehend language. Coupling this new knowledge with the knowledge derived from the comprehension systems now being developed and tested in the field of Artificial Intelligence will perhaps mean that a more complete model of human short story comprehension will indeed be realized in the not too distant future.

## BIBLIOGRAPHY

- (1) Kintsch, Walter, Memory and Cognition, John Wiley & Sons, Inc., 1977.
- (2) Kintsch, Walter, Representation of Meaning in Knowledge and Memory, Lawrence Erlbaum Associates, 1974.
- (3) Bobrow, D. and Allan Collins, Representation and Understanding, Academic Press, Inc., 1975.
- (4) Anderson, John R., Cognitive Psychology and Its Implications, W. H. Freeman and Company, 1980.
- (5) Nichols, Mary: KSU Interview topic: "Storytelling, the verbal art form", Nov. 29, 1982.
- (6) Lachman, R., Lachman, J. L., E. C. Butterfield, Cognitive Psychology and Information Processing: An Introduction, Lawrence Erlbaum Associates, 1979.
- (7) Schank, R. C., R. P. Abelson, Scripts, Plans, Goals and Understanding an Inquiry into Human Knowledge Structures, Lawrence Erlbaum Associates, 1977.
- (8) Sanford, A. J. and S. C. Garrod, Understanding Written Language Explorations of Comprehension Beyond the Sentence, John Wiley & Sons, Ltd., 1981.
- (9) Paivio, A. and I. Begg, Psychology of Language, Prentice-Hall, Inc., 1981.
- (10) Rumelhart, D. E., Introduction to Human Information Processing, John Wiley & Sons, Inc., 1977.
- (11) Avron Barr and Edward Fergenbaum (editors), The Handbook of Artificial Intelligence, Vol. 1, William Kaufmann, Inc., 1981.
- (12) Roger C. Schank and Christopher K. Riesbeck (editors), Inside Computer Understanding: Five Programs Plus Minutures, Lawrence Erlbaum Associates, 1981.

MODELING THE PROCESSES HUMANS USE TO  
COMPREHEND WRITTEN SHORT STORIES

by

SHERNISE ALEXA SPEARMAN

B.S., Kansas State University, 1980

---

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Computer Science

KANSAS STATE UNIVERSITY  
Manhattan, Kansas

1983

This report is a discussion of some of the current theories about and related to human written story comprehension. In particular theories and research from the fields of Artificial Intelligence, Cognitive Psychology and Psycholinguistics are emphasized. Although story comprehension involves several stages of processing only the more conceptual-level processes are detailed.

The first part of the report will provide the reader with a brief introduction to the general concept of human language comprehension, what it entails and how it is being examined and explained in terms of current comprehension theories. The rest of the report, sections one through four, shows how these theories and those from the field of Artificial Intelligence combine to produce computer comprehension systems.

Section one covers the analysis stage of the problem. First, the nature of understanding is questioned with respect to ways in which it can be tested. The idea of "levels of understanding" is examined introducing various knowledge structures and showing how they help organize the story to facilitate the understanding process. The emphasis of this is more on the human viewpoint than on the computer model of comprehension. Understanding the events of the story is also important; discussion of that topic introduces the concepts of event structures and predictability of events. This first section is designed to give the theories of several experts in the area of story comprehension which can then form the framework of knowledge needed to describe the computer models of story comprehension in the next section.

Section two discusses the representation of the text in the computer as well as speculates on the nature of the basic components of a theoretical understanding system. A general description of a system which



combines two current computer models that are currently working into one more complete hybrid computer comprehension system is outlined.

The third section is purposely short since the hybrid system above is merely speculation. In this section reference is made to the two currently operating systems mentioned in section two.

The final section addresses the question of how to test the running comprehension system for accuracy (i.e., how well it "understood" the story). Various methods are introduced, but not detailed (references are cited where more detailed accounts are given).

The report concludes with a brief summary message about the current state of affairs with respect to the theories of human story comprehension and the design of computer comprehension systems that attempt to model it.