

Perseverers, Recencies and Deferrers: New experimental evidence for multiple inference strategies in understanding

Richard H. Granger
Jennifer K. Holbrook

*Artificial Intelligence Project
Computer Science Department
University of California
Irvine, California 92717*

In the course of understanding a text, a succession of decision points arise at which readers are faced with the task of choosing among alternative possible interpretations of that text. We present new experimental evidence that different readers use different *inference strategies* to guide their inference behavior during understanding. The choices available to an understander range from various alternative inferential paths to the option of making no inference at a particular point, leaving a 'loose end'. Different inference strategies result in observably different behaviors during understanding, including consistent differences in reading times, and different interpretations of a text. The preliminary experimental results given here so far consistently support a previously published set of hypotheses about the inference process that we have called Judgmental Inference theory.

1.0 Introduction

When trying to understand even a simple text, readers make complex evaluations of the text to help choose one of several alternative interpretations. In making such decisions, readers employ a number of different strategies, including the selection of an *inference path*. This paper presents a theory of how readers make such choices. For example, subjects in our experiments read the following story:

[1] Melissa began to cry. Tyler had just asked her to marry him.

When asked why Melissa began to cry, different subjects gave at least two significantly different answers: (1) because she was upset for some reason about Tyler's proposal, perhaps because she couldn't or didn't want to accept the proposal; vs. (2) because she was so happy about the proposal that she was crying tears of joy. Preliminary findings in a series of controlled experiments indicate that

1. the answers a subject gives to this and similar questions correlate with different reading times;
2. these differences in reading times, and the differences in question-answering behavior, and the correlation between them, are all accurately predicted on the basis of hypothesized 'inference strategies' contained in Judgmental Inference theory [Granger 1982];
3. the different reading times and interpretations are not due to arbitrary individual differences, but rather to the adoption by

individuals of some particular principled inference strategy, as evidenced by the fact that individuals initially exhibiting one type of strategy can be experimentally manipulated to exhibit a different type of strategy.

While many theories of inference in understanding acknowledge the existence of alternative inferential paths, we present a theory that (1) catalogs the paths available to an understander, (2) predicts what mechanisms will lead to the choice of particular inferential paths during understanding, and (3) provides preliminary experimental evidence which supports these hypotheses.

2.0 Background

Language understanding is an interactive process which requires both adequate information presentation on the part of the communicator and skills of interpretation on the part of an understander. The interpretation skills necessary for understanding include mechanisms for such inference tasks as associating referents (Anderson & Bower, 1973; Haviland & Clarke, 1974), recognizing temporally or causally related events (Schank & Abelson, 1977; Bower, Black, & Turner, 1979; Black and Bern, 1980), and filling in unstated actions and preconditions for actions in a stereotypical sequence of events (Schank, 1975, 1977; Bower et al., 1979).

For example, consider the following story:

[2] Gail had accidentally poured coffee onto Will's hand. He screamed bloody murder.

Readers have no trouble inferring that "he" in the second sentence refers to Will from the first sentence. This is known as *referential cohesion*. Readers also recognize that Will screamed after the coffee was poured on him, although no specific temporal connection was stated. Further, readers will only infer that Will screamed because he was in pain or because he was very angry with Gail (or both). They will not infer that he screamed because he saw a ghost, or because he suddenly remembered it was time for his primal scream therapy. This connection between the two events is known as *causal cohesion* (Schank, 1977). Causal cohesion has been demonstrated extensively through various experiments (Anderson et al, 1973; de Villiers, 1974; Bower et al, 1979; Black et al, 1980), which indicate that recall of story events was higher and reading time was faster if the events in the story were causally related. These studies demonstrate the integral role inferences play in understanding text.

It is important to note that understanders make inferences about text as they read, not after they have finished reading. Referential inferences have been demonstrated to be made during reading. Seifert, Robertson, and Black (1982) discuss Haviland and Clark's 1974 experiments which show that reading times are longer when statements cannot be easily connected through referents. Seifert

et. al. also point out that inferences which do not rely on linguistic context are made by the reader. These *pragmatic inferences* involve causal connections formed between statements in the text and world knowledge which a reader draws upon to infer causality, case relations, missing events in a stereotypical sequence of events, and other abstract inferences. For illustration, recall story [2]. At the end of the first sentence an inference is made that the subsequent events will be related to the coffee spilling. This is easily illustrated by giving a different version of the story to readers:

[2a] Gail had accidentally poured coffee onto Will's hand. He told her a silly joke.

Readers would be confused by this version because there is no apparent causal connection. They would try to relate the events in any way possible; for example, they might decide that Will told the joke to Gail before she poured the coffee, and that she found the joke so funny that she got careless with her coffee pouring. A reader might also decide that telling jokes was Will's stoic response to pain, or that the coffee wasn't hot. Most readers will eventually conclude that the events were not causally related at all. What is intriguing is that any original expectations about events which a reader generates after reading the first sentence will be different from interpretations possible after reading the second sentence. There must be at least two points during the reading at which inferences about events can be made, for instance, one after the first sentence, and one after the second. Such points are known as *inference points* (see Rumelhart, 1981).

Several options are available to the reader at an inference point. The reader may leave a *loose end* (Granger, 1980a, 1981); that is, no particular inference about further events is generated. The reader often comes up with default inferences, which, given the material already read, are the most likely of the possible events, reactions, and so on. In story [2], some of the default inferences would be that the coffee had scalded Will, that he was in pain, and that he might react to the pain. The reader can also make another kind of inference which would not be based on the most likely outcome of the story based on the events so far. For example, in story [2], a plausible inference is that Will may decide to take revenge on Gail, or that Will screamed before Gail spilled the coffee, and she spilled the coffee because the scream startled her. Intuitively, the default inference about temporal relations is that events are described in the order in which they happened. Likewise, the most frequent default inference about causal relations is that earlier events caused subsequent events to happen.

However, it is not always the case that the order in which events are relayed in text is the order in which the reader infers them to

have happened. It only appears this way because texts are often written in a problem/solution fashion, rather than an action/explanation fashion. There are many texts that do not make it immediately clear whether or not events have been relayed in cause/effect order. When this happens, different strategies may arrive at different interpretations of the order.

The inferences that are made conform to *evaluation metrics* (Granger, 1980a) which determine how appropriate the inference is. Cohesion, in its various forms, is one such evaluation metric. Another metric is *parsimony* (Granger, 1980a, 1980b). Parsimony refers to the observation that readers will infer the least complicated explanation of events possible. For example:

[3] Mary picked up a magazine. She swatted a fly.

In the above story, readers will always infer that Mary picked up the magazine in order to swat the fly, even though several other interpretations are possible (for example, that she picked up the magazine to read, then was annoyed by the fly and used the magazine to swat it. Readers will agree that these more complicated interpretations are equally possible interpretations. But even though these interpretations are all equally logically possible, people universally choose the most *parsimonious* of the interpretations -- i.e., the one that will explain most parsimoniously the maximum number of events in the text.

If the reader left a loose end after reading the first sentence in the story above, as in figure 1a, there would be no problem relating the two sentences with the most parsimonious inference. However, if the reader makes the default inference that Mary wanted to read the magazine, as in figure 1b, the reader must *supplant* that initial inference with the new inference that the magazine was going to be used for flyswatting. Supplanting occurs when a default inference made by a reader does not account parsimoniously for all of the events in the text.

Readers are constantly evaluating their own inferences, supplanting inferences which are not parsimonious, re-interpreting whole sections of text when the text has a surprise ending (e.g. mystery stories, jokes), and so on. Judgmental Inference theory [Granger 1982] hypothesizes that all inference points are in fact decision points, where one of many paths is chosen (e.g. leaving a loose end, making a default inference, supplanting an inference, etc.) based on complex evaluations of the interpretation being constructed. These interpretations can involve a series of intermediate steps, in which many interpretations are tested. Judgmental inference is not a conscious attempt to discover the interpretation that is most parsimonious, most logical, or most cohesive. Rather, it is an unconscious process, guided by inference strategies.

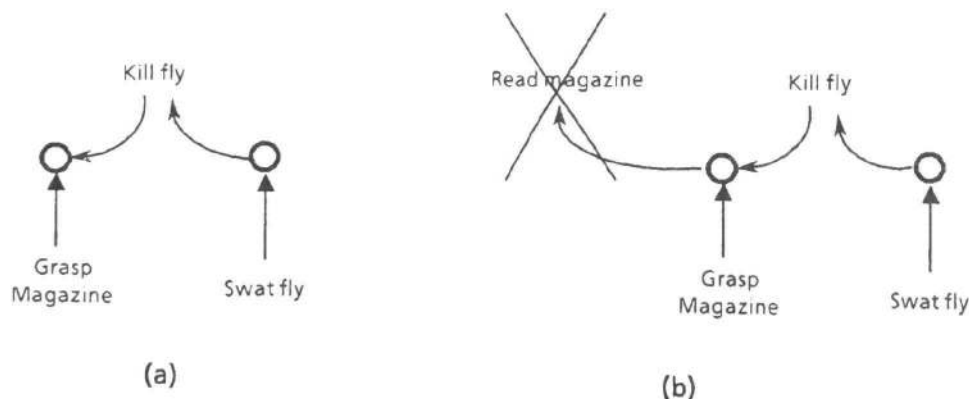


Figure 1

3.0 Theoretical Predictions of Judgmental Inference

Judgmental inference theory makes several specific predictions of how people choose 'inference paths'. This paper and the experiments described here focus on the study of these predictions. In particular, the experiments discussed below focus on the following questions:

1. When in the interpretation process are inferences made, if at all?
2. When in the interpretation process are loose ends left, if at all?
3. When do inferences get supplanted, if at all?
4. What happens when a reader doubts an inference?
5. When there are two (or more) equally plausible, parsimonious, and "normal" interpretations of story events possible, which one will readers select, why will they select it, and what will happen if doubt about the chosen interpretation is introduced?

It was pointed out above that a reader may take one of several inference paths when interpreting a story. For example, readers can leave loose ends or make inferences at inference points. Readers can read with a "naive" or a "suspicious" understanding: they can assume that the author is conveying the facts or trying to deceive them, that actors in the text have overt or covert goals, and so on. Yet most readers come up with very similar interpretations of story events. Either all readers follow the same inference paths, or with enough constraints, all inference paths will lead to similar interpretations of events. This theory espouses the latter view.

We theorize that interpretations are based on strategies or systematic choices between inference paths, and that individuals tend to use the same inference paths consistently. This would make possible the discovery of inference path choices, and explain why readers' interpretations of text are usually similar, yet diverge on occasion. However, the inference path system chosen is neither idiosyncratic nor universal to all readers. Instead, there seems to be a 'scale' of systems. We have divided this scale into several broad categories, although it should be noted that in reality, we have so far found no clear divisions among them.

Those readers who tend to come up with an interpretation of events as early as possible in the reading and then cling to that interpretation as long as possible can be described as *Perseverers*. Those readers who tend to leave loose ends about goals and plans (unless the goals or plans are explicitly stated) are known as

Recencies. Recencies will come up with an interpretation fairly late in the text, and if a conflict develops between possible interpretations, the interpretation based on the most recent information is chosen.

There are examples of extreme behavior at both ends of the scale. The most dedicated of Recencies will not make inferences. Unless a goal or a plan is explicitly stated, these readers will leave loose ends. Such behavior should result in quick reading, but slow and possibly haphazard answering when queried about inferred events. The extreme version of the Perseverer might be analogous to a "paranoid" reader. Such readers would make inferences based on preconceived notions. They might relate text to their own experiences, or ascribe attributes to characters which are clearly not in line with actual text.

Because there is not necessarily a clear division between the two main categories, there are readers who behave as though they could belong in either category. Such readers are called *Deferrers*. It is not clear whether Deferrers are using some combination of the other strategies or a different sort of strategy altogether. Future experiments may help explain Deferrer behavior.

Recall the following example:

- [1] Melissa began to cry. Tyler had just asked her to marry him.
Q: Why did Melissa cry?

As Figure 2a shows, a Perseverer reading this story will believe that Melissa was upset with Tyler's proposal; maybe she doesn't like Tyler, or maybe she is unable to get married even though she loves Tyler. The exact cause is unknown, but the default inference based on her tears is that she is unhappy. That is, in the absence of more specific information, crying is assumed to be a visible sign of pain or unhappiness, and there is some action which can explain unhappiness. Any new information in the story (here, Tyler's proposal) will be interpreted as an explanation for that unhappiness or a reaction to the unhappiness.

A Recency would believe just the opposite: that Melissa is happy with the proposal, and that she is crying tears of joy, not tears of sorrow. Figure 2b illustrates the inference strategy of a Recency trying to understand the events in this story. Recencies do not make an initial inference about Melissa's crying, not even a default inference; unless a specific goal or plan is stated, a Recency will leave a *loose end*, waiting for more specific information. If later events in the story are more specific about goals and plans, then earlier events will be interpreted in light of this more specific information. Even if no more specific information is given, a

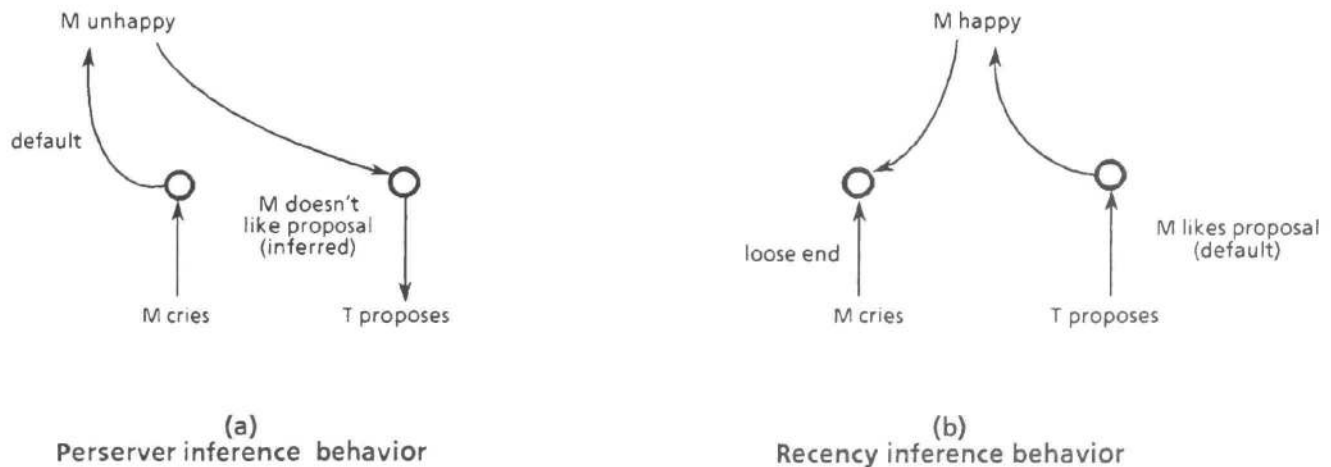


Figure 2

Recency will use the latest events in the story to interpret earlier events. Thus, a Recency would make no inference about the cause of Melissa's tears. The later information about Tyler's proposal would give rise to a presumed default interpretation of "happy event", and the earlier events in the story would be interpreted as being in line with a happy event, so that Melissa's tears are assumed to be tears of joy.

If this model is correct, reversal of the sentences of the above story should produce the opposite interpretation from the two extreme groups; that is, when presented with:

[4] Tyler had just asked Melissa to marry him. She began to cry.

Recencies should infer that Melissa is unhappy about the proposal, while Perseverers should infer that she is crying tears of joy.

4.0 Puzzles Solved

The theory of interpretation strategies helps answer the questions posed earlier. We can explain when loose ends will be left and when inferences will be made as being dependent upon which inference path the individual chooses. This explanation also suffices for predicting when inferences will be supplanted. Another problem that this model addresses is determining which of two equally plausible and parsimonious interpretations will be selected. Along these lines, Schank, et. al. explain such misunderstanding in verbal communication by "...maintaining that deriving a point is a part of processing, specifically related to the choice of an 'inference path'. Understanders choose to process idiosyncratically" (Schank, et.al., 1982, p. 263). This explanation of deriving a point agrees with our theory of inference paths. However, rather than believing understanders' processing to be idiosyncratic, this model predicts that individuals will tend to follow a single strategy consistently, rather than arbitrarily switching from path to path.

Another puzzle is presented in Rumelhart's (1981) work. His subjects had stories presented either a word at a time, a line at a time, or all at once. The subjects' inferences were collected either at the end of a line or at the end of the story. Rumelhart compared final interpretations of subjects who read the whole story, and subjects who read the story a line at a time, and wrote:

The results showed that subjects who interpreted a line at a time nearly always generated the same interpretations as those who gave us an after-the-fact interpretation. The only discernable difference was that those who gave an interpretation only at the end showed somewhat more variability in their interpretations (p. 27).

Rumelhart's own explanation of this phenomenon attempted to write it off as 'carelessness' on the part of the subjects:

It appears that this results from more careless reading on the part of the subjects offering an interpretation only at the end (p. 27).

However, viewed in terms of the inference strategies of Judgmental Inference theory, it is possible to interpret Rumelhart's data as further evidence for the hypothesized strategy paths. When text is presented to subjects a line at a time, with inferences about each line required, subjects are forced to act like Perseverers. Even if no inferences were elicited after each line, other demand characteristics of the task virtually force the subject to interpret the text in a particular manner. For example, if the text is presented a single line at a time, visual cues which would allow the subjects to recognize that there is more text available which might guide inferences would be lost. When subjects have all the text presented at once, they are free to interpret text using their usual strategy paths. Thus, the greater variability of

interpretation is an artifact of these different strategy paths, not a result of careless reading.

4.1 Experimental Validation of Hypothesis

Experiments are being conducted to discover whether the inference strategies exist, what characteristics should be ascribed to them, and whether individuals tend to use only one of the inference paths. In general, our experimental methodology is similar to that which Seifert, et. al. (1982) used in their experiments on pragmatic inferences. Our experiments, like theirs, utilize the methods of false recognition of material not found in the text, timing subjects' reading speed, and inquiring about the subjects' inferences only after a full text is read. The main difference was that Seifert, et. al., used texts of 17 lines each. They reasoned that it was possible that when readers were presented with only two lines of text, their inference strategies may be different than when reading a longer text; i.e., they may see it as only a story fragment, whereas a longer text looks like a full story.

The stories used in our experiment are not as long. However, we have controlled for the possible 'isolation effect' of short texts. The control stories used in this experiment varied in length: some of the stories were as short as the diagnostic stories, while some were several lines longer. If there is a difference in processing found between the long and short control stories, then it is likely that the 'isolation effect' is taking place in subject's analyses of the experimental texts. These techniques should make our experimental results externally valid.

4.2 Materials

Ten story sets, each consisting of one story and between six and nine questions, are presented to each subject as a single trial. Each story described a fairly stereotypical situation found in literature and the media. There were two kinds of questions to be answered. The first type of question required the subject to provide either information given in the story or an inference about the situation described in the story. The second type of question required the subject to make a truth judgment about the information in the question, which was either about information from the story or about inferences that could be made about events in the story.

Each trial had five control and five experimental story sets. Control stories were written to virtually force one shape of interpretation about the story situation; they were worded so that inferences would be made at the same points by everyone. Experimental, or *diagnostic*, stories were worded so that different shapes of interpretations are possible, and so that inferences need not be made at the same points by all readers, depending upon how the reader processes the story. Usually, the diagnostic stories allowed two nearly opposite shapes of interpretation. Also, the sentence order of experimental stories would permit rearrangement with the same shapes of interpretation possible, whereas the control stories' sentences could not be rearranged without destroying their sense.

Four versions of each diagnostic story were used. Some versions were permutations of sentence order, as explained above. Others had additional information which forced the shape of interpretation, but still allowed inferences to be made at different places in the text by readers using different. Because some interpretation shapes may tend to be more common than others, or interpretation shapes may be applied in a particular order, the different versions of the diagnostic stories had default inferences corresponding to several of the interpretation shapes. Thus, differences in subjects' interpretations could be accounted for by different processing methods, rather than particular interpretation shape biases. Only one version of each story appeared in a trial, so four trials with the different versions were constructed.

In addition to the different versions of the story, there were three *methods of presentation*. A story was either presented in its entirety or one sentence at a time, to test Rumelhart's results, discussed above. Asterisks (*) were in the text at inference points and at the end of every sentence. However, when the stories were presented a sentence at a time, the method of presentation was either with asterisks at inference points and at the ends of sentences, or only at the ends of sentences. Thus, there were three methods of presentation possible of the four versions of the story sets, for a total of twelve trials.

4.3 Procedure

Subjects were run individually. The subjects read instructions from an Apple II microcomputer, which informed them that they would take a reading comprehension test. The subjects were told to read the stories for comprehension rather than speed. They were to press the return key as soon as they read past an asterisk, either in the text or after a question. Subjects were instructed that all questions were to be answered; responses such as "I don't know", or "the story didn't say" were prohibited. Subjects were encouraged to answer with their best guess if they weren't sure of the correct response. It was suggested that the subject think of the stories as "situations", rather than respond with the actual text of the story.

No task intervened between a story and the questions. Reading time was recorded at each asterisk, which were placed at inference points, ends of sentences, and ends of questions. One question in each story set could be (randomly) re-presented, following the statement "That's a good answer, but there is a better one. Can you think of it?". Both answers to these questions would be recorded, as well as the order in which the questions were presented. A subject could be re-questioned from zero to five times, the number chosen randomly. Each subject was given as much time as necessary to complete the trial.

4.4 Results

Not all the data for this experiment have been collected yet. However, preliminary results indicate that the Recency and Perseverer strategy paths do exist. The theorized characteristics of both groups are also seem to be supported. In particular, Recencies leave loose ends when no goal or plan is stated, and make inferences consistent with the default inference of the latest text, as evidenced so far by reading times and question-answering data. Furthermore, Perseverers make initial inferences about goals and plans, and cling to the initial inferences whenever feasible. It also appears that each individual tends to favor a single inference strategy.

When text is presented to subjects a single line at a time, the final interpretations tend to be more uniform than when the text is presented all at once. Though it is still unknown if this effect is a significant one, this confirms Rumelhart's (1981) findings, and is consistent with the theory that without cues about text length, subjects are forced to act as Perseverers. This is evidence that strategy paths can be chosen by readers, and hence that different choices of inference paths are *not* due simply to individual differences, but are the result of distinct strategies.

5.0 Conclusion: Future Work

The experiments described here were only designed to confirm that the strategy paths exist by demonstrating that readers using the different strategies have different reading and understanding behavior, most notably, completely different interpretations of particular texts. However, these experiments did not carefully explore the underlying rules each strategy has which govern inference decisions. Other experiments are currently being designed to test hypotheses about the nature of these rules.

We have constructed a prototype for a computer program which models the Perseverer and Recency inference strategies, called STRATEGIST, described in Granger, Eiselt, & Holbrook (1983). STRATEGIST was based on the data we have collected from the experiments discussed in this paper. We intend to extend the STRATEGIST model and use it as a test-bed for hypotheses about strategy-driven inference rules, as well as exploring the role of the inference strategies with texts of different genres.

Acknowledgments

We acknowledge Dr. Art Graesser of California State University at Fullerton, and Dr. Ed Matthei of University of California at Irvine for many useful theoretical discussions and assistance in designing the experiments reported here. Amnon Meyers wrote the programs used to run the experiments.

References

- Anderson, J.R. & Bower, G.H. *Human associative memory*. Washington, D.C.: Winston and Sons, 1973.
- Black, J.B. & Bern, H. Causal coherence and memory for events in narratives. *Journal of Verbal Learning and Verbal Behavior*, 1981, 20, 267-275.
- Bower, G.H., Black, J.B. & Turner, T.J. Scripts in memory for text. *Cognitive Psychology*, 1979, 1, 177-220.
- de Villiers, P.A. Imagery and theme in recall of connected discourse. *Journal of Experimental Psychology*, 1974, 103, 263-268.
- Granger, R.H. *Adaptive Understanding: Correcting Erroneous Inferences* (Report No. 171). New Haven, Conn: Department of Computer Science, Yale University, 1980. (a)
- Granger, R.H. When Expectation Fails: Towards a self-correcting inference system. *Proceedings of the First National Conference on Artificial Intelligence*, Stanford, California, 1980. (b)
- Granger, R.H. Directing and Re-Directing Inference Pursuit: Extra-Textual Influences on Text Interpretation. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence* (IJCAI), Vancouver, British Columbia, 1981.
- Granger, R.H. Judgmental Inference: A theory of Inferential Decision-Making During Understanding. *Proceedings of the Fourth Annual Conference of the Cognitive Science Society*, Ann Arbor, Michigan, 1982.
- Granger, R.H., Eiselt, K., & Holbrook, J.H. *STRATEGIST: A program that models content-driven and strategy driven behavior* (Report No. 198). Irvine, CA: Artificial Intelligence Project, University of California at Irvine, 1983.
- Haviland, S.E. & Clark, H.H. What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 515-521.
- Rumelhart, D.E. *Understanding understanding* (Report No. 100). San Diego, Calif: Center for Human Information Processing, University of California at San Diego, 1981.
- Schank, R.C. & Abelson, R.P. *Scripts, Plans, Goals, and Understanding*. Hillsdale, N.J.: Erlbaum, 1977.
- Schank, R.C., Collins, G.C., Davis, E., Johnson, P.N., Lytinen, S., & Reiser, B.J. What's the Point? *Cognitive Science*, 1982, 6, 255-275.
- Seifert, C.M., Robertson, S.P. & Black, J.B. On-Line processing of pragmatic inferences (Report No. 15). New Haven, Conn: Cognitive Science Program, Yale University, 1982.

