Title
Statistical Detection of Local Coherence Relations in Narrative Recall and Summarization Data

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 25(25)

ISSN
1069-7977

Authors
Jaynes, Cynthia M.
Golden, Richard M.

Publication Date
2003

Peer reviewed

Statistical Detection of Local Coherence Relations in Narrative Recall and Summarization Data

Cynthia M. Jaynes (jaynes@utdallas.edu)[1]
School of Human Development, GR4.1
University of Texas at Dallas, Box 830688
Richardson, TX 75083-0688

Richard M. Golden (golden@utdallas.edu)[1]
School of Human Development, GR4.1
University of Texas at Dallas, Box 830688
Richardson, TX 75083-0688

Abstract

A new categorical time-series analysis method (which has a connectionist model interpretation) called KDC (Knowledge Digraph Contribution) analysis was used to investigate differences in recall and summarization production data as a function of reproductive and semantic coherence relations. The results provided support for the hypothesis that reproductive memory contributions play a dominant role in characterizing differences between recall and summarization. Moreover, the methodology and results described here illustrate the usage and application of KDC analysis.

Text Coherence and Comprehension

Semantic coherence relations, which indicate how specific clauses in a text are semantically related, play a major role in theories of human text comprehension (e.g., Kintsch, 1998), automatic text summarization (e.g., Mani, 2001; Mani and Maybury, 1991), and text linguistics (e.g., Longacre, 1979; Mann and Thompson, 1988) . In addition, considerable research in the field of experimental psychology has established that the organization of propositional information in the text (i.e., "textbase coherence relations") influences memory for text (e.g., Britton et al., 1980; Einstein et al., 1984; Golden, 1998; Hasher & Griffin, 1978; Kintsch & van Dijk, 1978; Kintsch, 1998). Moreover, it has been established that semantic coherence relations among ideas referenced within a text may be revealed through recall and summarization production data (e.g., Golden, 1998; Goldman & Varnhagen, 1986; Mandler & DeForest, 1979; Rumelhart, 1977; Stein & Glenn, 1979; Trabasso & Magliano, 1996). These observations suggest that it would be advantageous to develop a statistical methodology for the detection of local semantic coherence relations in human production data. Such a statistical methodology could be used not only for testing theories of human text comprehension but also for obtaining useful information for the purposes of informing the design of automatic summarization systems and refining theories of text linguistics. Towards this end, this paper presents an extension of work previously presented in Golden (1998) and applies the extended statistical modeling methodology to the analysis of some recall and summarization data which we have collected for the purposes of exploring differences and similarities between recall and summarization protocol data.

The Behavioral Experiment

Participants

Participants for this experiment were 24 undergraduate psychology students at University of Texas at Dallas who received research credit for participation. All participants were fluent English speaking students. Participants were randomly assigned to six conditions. Four participants were in each condition. The six conditions were comprised of different presentation orders of the three instruction conditions, recall, detailed summary and concise summary.

Table 1: The Czar and his Daughters
[Reprinted from Rumelhart, 1977].

"There was once a Czar who had three lovely daughters. One day, the three daughters went walking in the woods. They were enjoying themselves so much that they forgot the time and stayed too long. A dragon kidnapped the three daughters. As they were being dragged off they called for help. Three heroes heard the cries and set off to rescue the daughters. The heroes came and fought the dragon. They defeated the dragon and rescued the maidens. The heroes returned the daughters safely to their palace. When the Czar heard of the rescue he rewarded the heroes handsomely."

Materials

Two simple narrative texts, "The Dog and his Shadow" and "The Czar and his Daughters", taken from Rumelhart (1977) were used as practice and experimental stimuli for the study respectively. The experimental text consisted of nine

[1] The order of the authors is arbitrary.

sentences and 16 complex propositions. The practice text was of similar length and complexity. The experimental text is provided in Table 1.

Procedure

Text Presentation. A HyperCard program presented the practice text "The Dog and his Shadow" followed by the experimental text "The Czar and His Daughters" one sentence at a time using a self-paced reading task. Participants were instructed to read the instructions carefully and told they would be asked questions about the stories at a later time period in the experiment. To view each sentence, participants used an arrow in the corner of the screen. When the arrow was clicked, the present sentence disappeared and the next sentence appeared. After the experimental text was presented, participants participated in an intervening distractor task which took several minutes to complete.

Production Task. In the production task, participants were asked to summarize, give a detailed summary, and recall the story. The order of the instruction tasks were counter balanced across participants. Participants in the "concise summary" condition were instructed to construct a story summary consisting of no more than three sentences. Participants in the "detailed summary" condition were instructed to construct a story summary consisting of at least three sentences. Participants in the "recall" condition were instructed to recall the story by explicitly recalling the exact wording of each sentence and recalling the sentences according to their original order of presentation. Participants typed their responses for both the practice text and the experimental text into the HyperCard program.

Coding of Protocol Data

The recall and summarization protocol data for each individual participant in the study was then coded as an ordered sequence of complex propositions with the aid of the AUTOCODER computer program (Durbin, Earwood, and Golden, 2000). This computer program may be downloaded for research purposes only by visiting the web site: `www.utdallas.edu/~golden/autocoder.`

Proposed Situation Model

The proposed situation model is assumed to consist of two components: (1) a "reproductive memory" representation of the order in which the ideas in the text were presented, and (2) a "semantic local coherence" representation. These two semantic representations are formally expressed as directed graphs or "knowledge digraphs" which depict relations among a set of complex propositions. The complex propositions used in the proposed situation model are listed in Table 2. They were derived from Rumelhart's (1977) analysis of the "Czar and the Daughters" text. The proposition node number 17 refers to the initial mental state of the participant which arises from a request for the participant to recall or summarize the text. The proposition node 18 refers to the final mental state of the participant which arises when the participant's response is completed.

Reproductive Memory Knowledge Digraph

The reproductive memory knowledge digraph is the component of the situation model which refers to the original sequence of ideas which were presented to the reader. A formal representation of this knowledge digraph is specified by the arcs in Figure 1 which depict the order in which the complex propositions in Table 2 are presented in the original text. The reproductive memory knowledge digraph is intended to instantiate the behavioral hypothesis that production order is influenced by the original ordering of propositions in the text.

Table 2: Complex Propositions Used in Simulation Study (C=Czar, P=Princesses, D=Dragon, H=Heros)

Node Id #	Complex Proposition
1	SETTING(INTRODUCE,C)
2	SETTING(POSSESS(C,P)
3	METHOD(GO(P,WOODS))
4	CONSEQUENCE(POSSESS(P,JOY))
5	CONSEQUENCE(REMAIN(P,IN-WOODS)
6	METHOD(CAPTURE(D,P))
7	METHOD(TRANSFER(D,P))
8	CONSEQUENCE(SCREAM(P))
9	EVENT(HEAR(H,P))
10	METHOD(GO,H)
11	SETTING(INTRODUCE,H)
12	METHOD(KILL(H,D))
13	METHOD(RESCUE(H,P))
14	METHOD(TRANSFER(H,P))
15	EVENT(HEAR(C,NEWS))
16	OUTCOME(REWARD(C,H))
17	INITIAL-CONTEXT
18	FINAL-CONTEXT

Semantic Coherence Knowledge Digraph Based Upon Rumelhart's (1977) Story Grammar

The semantic coherence knowledge digraph is intended to model the predictions of Rumelhart's (1977) theory for predicting the order in which experimental participants will generate sequences of propositions when they were asked to summarize the experimental text. According to Rumelhart's semantic analysis of the experimental text, episodes in the story are hierarchically organized with the "lower-level" episodes describing the achievement of the subgoals for the "higher-level" episodes. Rumelhart's (1977) model makes predictions about the levels of summarization for this story. Rumelhart's (1977) predictions regarding Level 0, Level 1 and Level 2 summaries for "The Czar and his Daughters" are summarized in Figure 2. A concise summary from Rumelhart's model would be modeled by the arcs in Figure 2 as the path: 17, 6, 13, 16, 18 While more detailed summaries would be modeled by other paths through the network presented in Figure 2 (e.g., 17, 6, 8, 13, 16, 18).

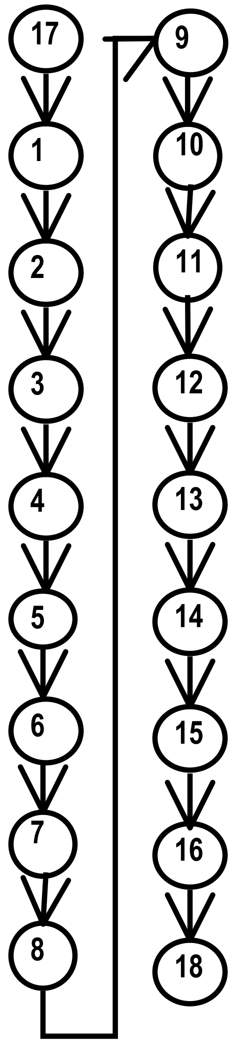


Figure 1: Reproductive Memory Knowledge Digraph. Each arc indicates the order in which a proposition in Table 1 follows another in the original text.

Knowledge Digraph Contribution Analysis

Overview

Knowledge Digraph Contribution (KDC) analysis is a special type of categorical time-series analysis which is specifically intended to identify evidence for different types of knowledge digraphs through the analysis of ordered sequences of propositions in production data. More specificially, KDC theory is based upon a specialized type of multinomial logistic regression time-series analysis where individual beta weights in the model correspond to the influence of different knowledge digraphs. . The KDC software used here may be downloaded (for research purposes only) from the web site: www.utdallas.edu/~golden/kdc. Earlier versions of KDC theory (e.g., Golden, 1994, 1995, 1998) based the sample size on the number of participants in the experiment instead of the number of propositions mentioned by the participants. By reworking the mathematics such that the sample size is based upon the number of propositions mentioned following the method of Golden (in press), the statistical power of the KDC theory is dramatically improved.

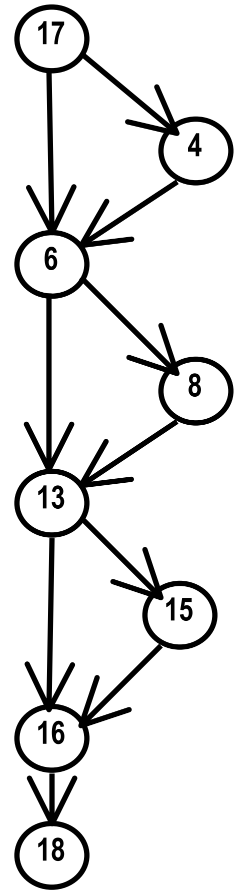


Figure 2: Rumelhart Summarization Knowledge Digraph. Each arc indicates the order in which a proposition in Table 1 is expected to follow another in a story summary based upon a story grammar causal network analysis.

Statistical Model

Formally, let $\mathbf{x}_i(t)$ indicate the *t*th proposition mentioned by the *i*th participant in the experiment within a particular experimental condition where the notation $\mathbf{x}_i(t)$ denotes a *d*-dimensional vector. If $\mathbf{x}_i(t)$ refers to the *k*th proposition in the proposition dictionary, then $\mathbf{x}_i(t)$ is the *k*th column of a *d*-dimensional identity matrix. In such a situation, the vector $\mathbf{x}(t)$ defined as the *k*th column of a *d*-dimensional identity matrix would denote that the *t*th proposition mentioned by a particular participant within a particular experimental condition was the *k*th proposition in a dictionary of *d* propositions. Let $p(\mathbf{x}(t)=\mathbf{u}_k)$ denote the probability that the the *t*th proposition mentioned by the participant will be the *k*th proposition in the proposition dictionary. Let $\mathbf{R}$ be a matrix which specifies the reproductive memory knowledge digraph in Figure 1 such that the *ij*th element of $\mathbf{R}$ is equal to one if the *j*th node in Figure 1 is connected to the *i*th node by an arc and a zero otherwise. Similarly, let **C** be a matrix which specifies the causal knowledge semantic coherence relation digraph in Figure 2 such that the the *i*th element of **C** is equal to one if the *j*th node in Figure 2 is connected to the *i*th node by an arc and a zero otherwise. Let β_R and β_C be the "contribution weights" indicating the respective influences of the *d*-dimensional square knowledge digraph matrices **R** and **C**. Then the specific KDC statistical model used here is specified by:

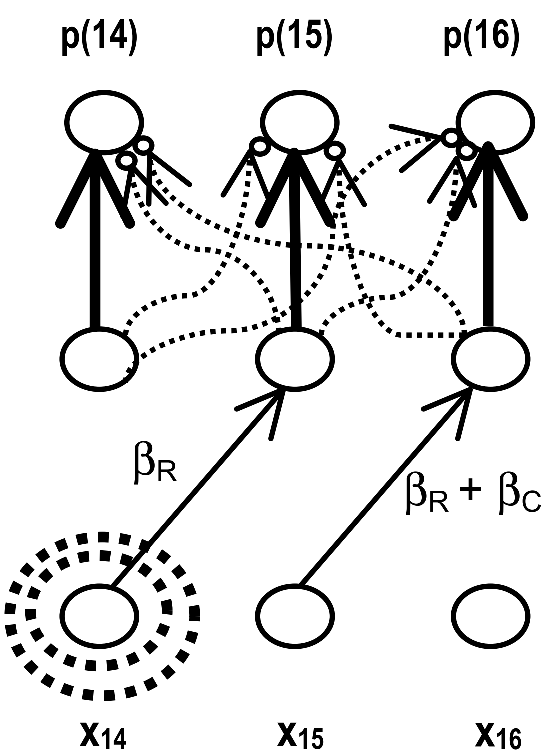
$$p(\mathbf{x}(t)=\mathbf{u}_k) = \exp((\mathbf{u}_k)^T\mathbf{h}_k)/\sum_m (\mathbf{u}_k)^T\mathbf{h}_m \qquad (1)$$

for $m = 1, \ldots, d$ and where:

$$\mathbf{h}_m = [\ \beta_R\ \mathbf{R} + \beta_C\mathbf{C}\ ]\ \mathbf{x}(t\text{-}1). \qquad (2)$$

Figure 3: Connectionist Interpretation of KDC Model. Only 3 of the 18 proposition nodes (propositions 14,15, and 16 in Table 1) are shown. In this example, the model has just mentioned proposition 14 in Table 1 and the probabilities that the model will generate propositions 14, 15, and 16 (denoted as p(14), p(15), and p(16)) are computed and returned as the output unit activations of the network.

The two free parameters in the model are the contribution weights β_R and β_C which indicate the respective predictive influence of knowledge digraphs **R** and **C**. Using the large sample methods of Golden (in press; also see White, 1994) maximum likelihood estimates of the contribution weights can be estimated in conjunction with their standard errors. In addition, statistical tests can be constructed for providing insights regarding how the beta weights change as a function of the instruction condition in the experiment.



Connectionist Model Interpretation

Finally, note that as shown in Golden (1998), the probabilistic model in KDC theory has a dual interpretation as a particular type of highly structured connectionist network. Figure 3 shows a portion of the connectionist network interpretation of Equations (1) and (2). The activation levels of the input units are denoted by x_{14}, x_{15}, and x_{16} and only x_{14} is activated indicating that proposition 14 was most recently mentioned by the model. The next layer of connection strengths which connect the input units to the hidden units are specified by the knowledge digraphs in Figure 1 and Figure 2. Only two parameters β_R and β_C are estimated in order to obtain the complete mapping from input units to hidden units. The activation levels of the hidden units are specified by Equation (2) in the text. The output layer of connection strengths is a forward lateral inhibition mapping which is constant and implements the "softmax" nonlinearity described by Bridle (1990). The output activations are probabilities which are always positive and sum to one and are specified by Equation (1) in the text.

Computing maximum likelihood estimates using KDC theory is formally equivalent to having the connectionist network *learn* to recall or summarize texts using human production data as training data. Production data can be *generated* from the connectionist network by sampling from the KDC probability model.

Parametric Bootstrapping

In order to check the validity of the large sample approximations, three such connectionist networks were constructed for each of the three experimental conditions. Then, each of these three networks was used to generate three simulated response data sets where each data set consisted of the responses from 24 simulated subjects. The generation algorithm consisted of simply sampling from the KDC probability model described by Equations (1) and (2) to generate three additional "simulated" data sets. Then for each simulated response data set the contribution weights were estimated. Since there were three parameter estimations for each of the three experimental conditions involving both the reproductive memory and semantic coherence contribution weights, 18 additional contribution weights were estimated. These additional 18 contribution weights will be referred to as "bootstrap contribution weight estimates". If the large sample approximations are correct, then the standard errors estimated by the theory should approximately contain the bootstrap contribution weights estimated/learned from the simulated data.

Results

Qualitative Overview

The results of applying KDC theory to the data in this study are reported in Figure 4. As can be seen from Figure 4, the influence of the Reproductive Memory Knowledge Digraph Contribution Weight β_R has the greatest influence when participants are asked to recall a text, a moderate influence for the detailed summary condition, and the least influence in the concise summary condition. In contrast, the influence of the Semantic Coherence Knowledge Digraph Contribution Weight β_C has the greatest influence in the concise summary condition and shows a tendency to decrease in the detailed summary condition and decrease further in the recall condition.

Statistical Test Results

Within-Group KDC Statistical Tests As previously noted in Figure 4, the β_R and β_C contribution weights were estimated using each of the three data sets yielding six beta weight estimates. KDC theory was then used to estimate the standard errors for each of the six beta weights. These standard errors are plotted as confidence intervals in Figure 2 as well. All six beta weight coefficients were significantly positive ($p < 0.0001$) indicating a statistically significant contribution of all knowledge digraphs.

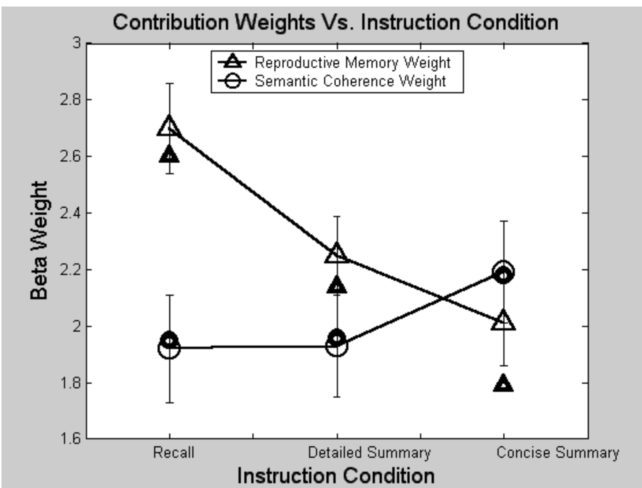


Figure 4: Reproductive Memory and Causal Knowledge Digraphs. Reproductive memory knowledge digraph is depicted by the solid arcs. Rumelhart's semantic local coherence causal knowledge digraph is depicted by the dashed arcs. Smaller circles and triangles denote bootstrap contribution weight estimates (see text for details).

Between Group KDC Statistical Tests As shown in Figure 4, a between-group planned comparison KDC statistical test revealed the β_R and β_C contribution weights estimated in the concise instruction condition differed reliably from the contribution weights estimated in the recall instruction condition, $W(2) = 10.3$, $p = 0.006$. Post-hoc analyses showed this reliable difference appeared to arise from a smaller reproductive memory contribution β_R weight ($Z=3.1$, $p = 0.002$) in the concise relative to the recall instruction conditions. Although the β_C weight was larger in the concise instruction relative to the recall condition as shown in Figure 2, this difference was not statistically significant ($Z=1.0$, $p = 0.3$).

A second KDC between-group planned comparison test showed only a marginally significant change in the pattern of beta weights between the detailed summary and recall instruction conditions, $W(2) = 4.7$, $p = 0.10$ (please see Figure 4). Post-hoc analyses showed this marginally significant difference appeared to arise from a smaller reproductive memory contribution β_R weight ($Z=2.1$, $p = 0.03$) in the detailed summary relative to the recall instruction conditions. Although the semantic coherence β_C weight was larger in the detailed summary instruction relative to the recall condition as shown in Figure 4, this difference was not statistically significant ($Z=0.04$, $p = 0.97$).

A third KDC between-group planned comparison test showed no significant change in the pattern of beta weights between the detailed summary and concise summary instruction conditions, $W(2) = 2.1$, $p = 0.35$ (despite the clear trends in the data depicted in Figure 4).

Reliability of the Asymptotic Statistical Inferences

The reliability of the estimated variance-covariance matrix of the parameter estimates is indirectly checked by seeing if the bootstrap contribution weights lie approximately within the estimated confidence intervals. Inspection of Figure 4 shows the bootstrap contribution weights (identified by the small circles and small triangles) tended to lie within the estimated confidence intervals with the sole exception of one bootstrap reproductive memory contribution weight estimate in the concise summary condition. These observations supports the appropriateness of the large sample approximations used in this paper.

General Discussion and Summary

Qualitative trends obtained from the KDC data analysis showed that semantic knowledge components of the situation model are more dominant than reproductive memory situation model components for summarization data. Moreover, the reverse pattern tends to hold for recall protocol data. KDC statistical analyses indicated that the semantic knowledge components of the situation model provide influence production data in a similar way in both recall and summarization. On the other hand, the reproductive memory knowledge digraph had a greater influence in the recall condition relative to the summarization conditions. Simulation studies supported the reliability of the large sample statistical inferences which only revealed a reliable difference in the pattern of beta weights between the concise-instruction and recall-instruction experimental conditions.

Because of the clear trends in the data shown in Figure 4, a likely interpretation of the obtained results is that the lack of reliable differences between the detailed summary instruction condition and the other recall and concise summary instruction conditions is simply due to a lack of statistical power associated with the KDC analysis. An alternative interpretation is that the lack of reliable differences might simply not be present indicating that differences in recall and summarization performance arise primarily due to reproductive memory processes.

In summary, a new categorical time-series analysis method called KDC analysis was applied to the analysis of previously unpublished recall and summarization protocol data. In addition to providing some new insights into the relationship between recall and summarization data, the methodology and results described in this paper provided an example of the usage of KDC analysis within the context of a practical cognitive modeling problem and illustrated how the large sample approximations could be checked using a connectionist model for parametric bootstrapping simulated data generation.

Acknowledgments

This research was supported in part by the Information Technology Research (ITR) Initiative through the Research on Learning and Education Program Award 0113369 to Richard M. Golden.

References

Bridle, J. S. (1990). Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In *Neuro-computing: Algorithms, Architectures, and Applications* (F. Fougelman-Soulie and J. Herault, eds.), pp. 227-236. New York: Springer-Verlag.

Britton, B., Meyer, B., Hodge, M., & Glynn, S. (1980). Effects of the organization of text on memory: Texts of retrieval and response criterion hypotheses. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 620-629.

Durbin, M. A., Earwood, J., & Golden, R. M. (2000). Hidden Markov models for coding story recall data. In *Proceedings of the 22nd Annual Cognitive Science Society Conference.* Mahwah, New Jersey.

Einstein, G., McDaniel, M., Bowers, C., & Stevens, D. (1984). Memory for prose: The influence of relational and proposition-specific processing. *Journal of Experimental Psychology: Learning Memory and Cognition, 10,* 133-143.

Golden, R. M. (1994). Analysis of categorical time-series text recall data using a connectionist model. *Journal of Biological Systems, 2, 283*-305.

Golden, R. M. (1995). Making correct statistical inferences using a wrong probability model. *Journal of Mathematical Psychology, 38*, 3-20.

Golden, R. M. (1998). Knowledge digraph contribution analysis of protocol data. *Discourse Processes, 25,* 179-210.

Golden, R. M. (in press). Discrepancy risk model selection test theory for comparing possibly misspecified or nonnested models. To appear in *Psychometrika.*

Goldman, S & Varnhagen, C. (1986). Memory for embedded and sequential story structures. *Journal of Memory and Language, 25,* 401-418.

Graesser, A. C., Robertson, S. P., & Anderson, P. A. (1981). Incorporating inferences in narrative representations: A study of how and why. *Cognitive Psychology, 13*, 1-26.

Hasher, L. & Griffin, M. (1978). Reconstructive and reproductive processes in memory. *Journal of Experimental Psychology: Learning Memory and Cognition, 4*, 318-330.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* New York: Cambridge University Press.

Mandler, J. M. & DeForest, M. (1979). Is there more than one way to recall a story? *Child Development, 50,* 886-889.

Mani, I. (2001). *Automatic summarization.* Philadelphia: John Benjamins Publishing Company.

Mani, I. and Maybury, M. T. (1999). *Advances in automatic text summarization.* Cambridge: MIT Press.

Mann, W. C. and Thompson, S. A. (1987). Rhetorical structure theory: Towards a functional theory of text organization. *Text, 8,* 243-281.

Kintsch, W. and van Dijk (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363-394.

Rumelhart, D. E. (1977). Understanding and summarizing brief stories. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension* (pp. 265-303). Hillsdale, NJ: Erlbaum.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing: Vol. 2. Advances in discourse processes* (pp. 53-120). Norwood, New Jersey: Ablex.

Trabasso, T. & Magliano, J. P. (1996). Conscious understanding during comprehension. *Discourse Processes, 21,* 255-287.

White, H. (1994). *Estimation, inference,and specification analysis.* New York: Cambridge University Press.