Are Seven Words All We Need?
Recognizing Genre at the Sub-Sentential Level

Philip M. McCarthy (pmccarthy@mail.psyc.memphis.edu)
Danielle S. McNamara (d.mcnamara@mail.psyc.memphis.edu)
Department of Psychology
Memphis. TN 38152

Abstract

Genre recognition is a critical facet of text comprehension. In this study, we assess the minimum number of words in a sentence necessary for genre recognition to occur. Using corpora of Narrative, History, and Science sentences, we found that three experts in discourse psychology (demonstrating high agreement) accurately recognized the genre of over 80% of the sentences. This recognition generally occurred within the first seven words, with the highest accuracy for the Narrative genre. Thus, even very short and incomplete text can potentially activate text-structure knowledge and facilitate comprehension. In addition, we show that Narrative-like sentences are the most pervasive sentence type, with expert raters assigning 51% of misclassified sentences to the Narrative genre (again with high agreement between raters). In contrast, only 11% of misclassified sentences were assigned to Science. This study allows us to establish baseline expectations for skilled readers so that we can further examine differences in speed and accuracy of genre recognition as a function of reading skill.

Keywords: Genre; genre recognition; domain; register; text; reading comprehension; narrative; expository.

Introduction

A reader's comprehension of a text can be facilitated by correctly identifying the textual characteristics that indicate its genre (Bhatia, 1997; Graesser, Olde, & Klettke, 2002). Indeed, knowledge of text structure is an important facet of reading skill, and training to recognize text structure helps to improve struggling readers' comprehension (Meyer & Wijekumar, 2007; Oakhill & Cain, 2007; Williams, 2007). Research indicates that skilled readers activate particular expectations and strategies depending on the genre to which they attribute the text. Comprehension and subsequent learning can be facilitated because these strategies assist in the encoding and retrieval of content from episodic long-term memory (van Dijk & Kintsch, 1983).

The goal of this study is to investigate experts' ability to recognize the genre of sentences presented out of context. The premise of the study is that speed and accuracy of recognizing the genre of sentences may provide a signature of reading ability, and thus the development of a task to do so may provide a method of assessing reading ability, or at least some aspects of reading ability. This study describes our work to create and better understand this task. Specifically, we examine here whether three experts in discourse psychology agree on the genre classification for isolated sentences (i.e., is the task even possible) and how many words are required for accurate genre classification. Beyond our immediate goal of establishing a novel reading assessment paradigm, this study also provides insights on the processes of text comprehension and the compositionality of genres.

Genres and Domains

The term *genre* designates a category of text (Graesser et al., 2002). At the highest taxonomic level, Narrative text is generally contrasted with Expository text (e.g., McDaniel et al., 1986). This distinction posits that the structure of Narrative is more easily mapped onto everyday experience and, as a result, readers tend to process the global and thematic relationships in a passage (Otero, Leon, & Graesser, 2002). In contrast, Expository texts are more likely to discuss unfamiliar topics. Consequently, the lack of sufficient prior knowledge forces higher ability readers to process the details of the text at a more local level (e.g., connections between adjacent clauses). Empirical evidence supports such theories through recall (Graesser et al., 1980) and reading time experiments (Graesser, Hoffman, & Clark, 1980) demonstrating that Narrative text is recalled approximately twice as accurately as Expository text, and also read approximately twice as fast.

At a lower taxonomic level, *domains* of Expository texts (such as History and Science) exhibit conventional features that are familiar to members of their relevant discourse community. These discourse features guide the readers' attention, comprehension, and memory (Graesser et al., 2002).

In this study, we consider three domains: Narrative, History, and Science. We include History because whereas no one disputes that Science texts are representative of the Expository genre, there is a question as to whether History is more Expository-like or more Narrative-like. Some researchers, for example, have recognized that History texts can be similar to Narratives, the two domains tending to be presented more as a chronological series of events on topics with which many readers are familiar (Tonjes, Ray, & Zintz, 1999). Empirical computational approaches to distinguishing the genres provide evidence for both categorizations: For instance, McCarthy, Graesser, and McNamara (2006) used an array of cohesion indices showing that History texts were more similar in structure to Science texts. That is, both History and Science texts were more cohesive than Narrative texts. On the other hand, Duran et al. (2007) used temporal indices and found evidence that History texts were more similar to Narratives. That is, both History and Narrative texts were structured

similarly in terms of temporal development. Meanwhile, Lightman et al. (2007) found evidence for all three domains having distinct characteristics. Thus, one question addressed in this study is whether History sentences are correctly classified to a similar degree as Narrative and Science sentences; and if not, to which domain are they more likely to be assigned.

Purpose of the Study

Research indicating the importance of genre recognition is substantial. However, the genre characteristics that led to this recognition are less well understood. More specifically, the amount of text required to correctly identify a genre has received little if any attention. As such, to better understand how language is recognized, we investigate at what point genre recognition takes place when reading a sentence. Such a study may offer significant implications for research in knowledge representation, reading ability development, and computational text analyses. As such, in this study, we focus on two research questions: 1) can experts identify the domain of texts at the sentence level, and 2) if so, how many words do experts need to make that identification?

Predictions

For the Narrative domain, we predicted that incorrectly assessed sentences would more likely be classified as History sentences because both domains typically describe past events. For the History domain, we predicted misclassified sentences to be equally distributed, because History texts are equally likely to be descriptive of an event (thus, Narrative-like) or feature explicit lexical cause and effect relationships (thus, Science-like). For the Science domain, we predicted that misclassified sentences would more likely be assessed as History sentences, because some elements of scientific texts present explanations from a chronological perspective.

We further predicted that our expert raters would correctly identify a high percentage of sentences requiring approximately only half of the words in a sentence to do so. This prediction is based on typical features of verb and pronoun positioning. Verbs, for example, feature early in a sentence, and their tense is indicative of their genre (McCarthy et al., 2007). Similarly, the subjects of sentences are generally positioned at the beginning of sentences. Regardless of whether the subject of the sentence is a pronoun or named entity, the characteristics of the sentence subject are at least somewhat indicative of text genre.

Method

Three researchers in discourse processing (one post-doc, one graduate student, and one advanced (published) undergraduate) assessed 210 sentences equally representing the domains of Narrative, History, and Science.

The corpus in our analysis was composed of a subset of sentences taken from the 150 academic text corpus compiled by Duran et al. (2007). In that corpus, the texts were sampled from 27 published textbooks provided by the MetaMetrics repository of electronic duplicates. A subset of the Duran and colleague's corpus (McCarthy et al., in press) further focused the corpus by filtering out an equal number of similarly sized paragraphs. The McCarthy and colleague's sub-corpus featured 207 paragraphs in total (828 sentences): 69 paragraphs in each of the three domains, and 23 paragraphs each of 3, 4, and 5 sentences in length. The approach we adopted for sentence selection from these paragraphs is based on studies indicating that topic sentences are processed differently to other sentences in a paragraph (e.g., Kieras, 1978, Clements, 1979, McCarthy et al., 2007). Because such research also indicates that topic sentences are more likely to occur in the paragraph initial position (Kieras, 1978, McCarthy et al., 2007), we sampled an equal number of paragraph-initial sentences and paragraph-non-initial sentences. For the paragraph-non-initial sentences, we used the third sentence of each paragraph. This choice was made for two reasons. First, all paragraphs contained a third sentence; and second, third-sentences are presumably less closely related in terms of co-reference to first-sentences than first-sentences are to second-sentences; thus, the effects of a possible confound are reduced. This reduction to first-sentences and third-sentences left 414 candidate sentences in our corpus. To ensure that participants viewed sentences of approximately equal length, we further reduced the size of the corpus by only including all sentences that were within one SD of the average length in terms of number of words of the 414 candidate sentences (mean number of words = 15.437; SD = 7.113). Using this criterion, 298 sentences remained, of which the smallest group was 35 sentences belonging to the domain of narrative-paragraph-non-initial. We thus selected 35 to be the number of sentences from each of the six groups (Narrative/History/Science by paragraph-initial/paragraph-non-initial). Consequently, our corpus consisted of 210 sentences, equally representing the three domains and the initial/non-initial sentence dichotomy.

Procedure

A Visual Basic program was created to evaluate genre recognition. The program included three parts: *instructions*, *practice examples*, and *testing*. Following the instructions, participants were provided with six practice sentences. Once the practice was completed, a message informed the participants that the experiment would begin. Each participant evaluated all 210 sentences. The sentence order was randomized for each participant. The program operated by displaying the first word of the first sentence in a text window. Participants were required to assess the domain to which they thought the sentence fragment belonged. Participants registered their choice by clicking on one of four on-screen buttons: *Narrative*, *History*, *Science*, and *Don't Know*. As soon as a genre choice was made, the next word from the sentence appeared in the text window. All punctuation was retained in the display and was attached to the word it adjoined (e.g., in the sentence fragment *Yes, it was a …* the word *Yes* would appear as *Yes* + comma).

Table 1: Accuracy and misclassifications for Narrative, History, and Science texts, and "Don't Know"(DK) classifications.

	Accuracy			Correct			Misclassification			
	Recall	Precision	F1	Narrative	History	Science	Narrative	History	Science	DK
Rater 1	.824	.840	.832	.914	.829	.729	.081	.052	.023	.019
Rater 2	.824	.892	.856	.871	.857	.743	.062	.038	.000	.076
Rater 3	.810	.817	.813	.886	.757	.786	.081	.076	.024	.029
Mean	.819	.850	.834	.890	.814	.752	.075	.055	.016	.041

After 10 seconds, if the participant made no decision, then a new word automatically appeared in the text window with a message informing the participant of the new word. The variables of *genre choice* and *accuracy* were recorded. Participants evaluated each word of each sentence until they had either given the same decision of the genre of the sentence three consecutive times (whether right or wrong), or until all the words in the sentence were presented. The final choice of participants was recorded as the genre choice, regardless of previous decisions.

Results

Raters

We begin our analyses by demonstrating inter-rater reliability. This reliability establishes confidence in our evaluation of the data as typical of expert ratings and is particularly important when using few raters. On average, our raters correctly identified the genre of the sentences for 90% of the data. Inter-rater agreement between Raters 1 and 2 for correctly assessed sentences was approximately 90% (X^2 = 41.077, $p < .001$). Inter-rater agreement between Raters 1 and 3 was also approximately 90% (X^2 = 47.569, $p < .001$). And the Inter-rater agreement between Raters 2 and 3 was approximately 91% (X^2 = 61.145, $p < .001$).

Of the 210 sentences assessed, *all three* raters classified the correct genre for approximately 69% of data. Two of the three raters correctly classified an additional 17% of the sentences. At least one of the three raters correctly identified an additional 6% of the data. Also, less than 9% of the data were incorrectly assessed by any of the raters. Thus, the raters' accuracy was quite high. Further reliability of the raters' analyses can be demonstrated in terms of recall and precision (see Table 1). Such accuracy and agreement between the three raters (M=82%) offers support for the forthcoming analyses to be considered representative of genre identification at the word level by experts in discourse processing.

Genre

In terms of genre recognition accuracy, the expert raters correctly classified 516 of the 630 sentences: an average accuracy of 82% (see Table 2). This result is in line with our prediction. While the results appear consistent across the genres (Min. F1 = 82, Max. F1 = 84), closer analyses suggest that the genres elicit quite distinct patterns of responses.

Narratives The Narrative domain received the highest recall value (89%); however the 47 additional false alarms made the Narrative domain the least precise (80%). Indeed, of all misclassifications, more sentences were incorrectly assigned by the experts as Narrative, than either of the two expository domains (Narrative = 51%; History = 38%; Science = 11%). The misclassifications to the Narrative domain suggest that Narrative sentence structures may be the most pervasive type. The approximately equal division of false alarm Narrative sentences to the Science (22) and History (25) domains further suggests that the two Expository domains may comprise, to a small but notable degree, Narrative-like sentences. Indeed, for six sentences (three History and three Science) all three-raters categorized the sentences as Narratives (see Table 3).

Table 2: Accuracy and misclassifications of expert raters by domain for Narrative, History, and Science texts, and unclassified "Don't Know" (DK) texts.

	Decisions		Accuracy			Misclassifications			
Domain	Selected	Correct	Recall	Precision	F1	Narrative	History	Science	DK
Narrative	234	187	0.890	0.799	0.842	/	10	3	10
History	206	171	0.814	0.830	0.822	25	/	7	7
Science	168	158	0.752	0.940	0.836	22	25	/	5

Table 3: The six sentences identified by all raters as narratives.

Example	Domain	Sentence
1	History	We cannot sell the[1] lives of men and animals, said[3] one[2] Blackfoot chief in the 1800s, "therefore we cannot sell this land."
2	History	I had vainly[1] flattered myself[3] that[2] without very much bloodshed it might be done.
3	History	Much to my surprise[1], I had[2] forgotten[3] my glasses in prison, so I used my wife's.
4	Science	Taking no joy in life[1], looking forward to nothing[3], wanting to withdraw from people and activities[2].
5	Science	This, he thought, would demonstrate[1] that[2] emotions[3] can be mechanically induced (Cohen, 1979).
6	Science	Watson went even[1, 3] further and[2] suggested that at the human level, deep emotions are also just the result of association and learning.

Note: The superscript number indicates the point at which the final genre selection was made

Looking more closely at these "misclassified" sentences, we observe that all three raters classified Example 1 as Narrative by the 11th word of the sentence. It is only after this point that the words *Blackfoot chief* reveals the sentence more clearly as a History text. For example 2, all three raters classified the text by the 6th word. Indeed, although the text recounts an historical event, the use of first person pronoun (rare in an expository structure) may be indicative of a Narrative style of writing. This appears again in example 3. All three raters classify the sentence in example 3 by the 7th word. Again, the incorporation of first-person pronouns renders the sentence more Narrative-like, even though the text as a whole is taken from a History book. Example 4 is actually a sentence fragment and resulted in one rater having to view the entire sentence before deciding that it was Narrative. While the sentence lists symptoms of depression, the text could easily be read as describing a character. For example 4, all raters agreed on Narrative by the 7th word. Had the raters read a little further, however, the Science-like nature of the sentence (passive construction) may have been more easily recognized. The final example is deemed Narrative by the fifth word. It is possible that the raters saw the subject word *Watson* and considered the text to be from Sherlock Holms. The results are in line with our predictions that the early presence of key lexical and grammatical features triggers the expert readers' genre recognition.

History As predicted, when History sentences were misclassified they tended to be identified as Narratives. This result supports the conclusions of Duran et al. (2007) and Tonjes et al. (1999). The three examples above (Table 3) demonstrate the type of Narrative-like text that appears to be a feature of History texts.

Science Only 75% of the Science sentences were classified accurately, the lowest of the three domains. However, when raters did label a sentence as Science they were nearly always correct to do so (precision = 94%, the highest of the three domains). Of the 52 misclassified science items, most were attributed to History (25) and Narratives (22). The high History value is as predicted, because much scientific discussion begins from a historical perspective. The equally high Narrative value suggests that Science texts may be equally viewed as Narrative-like in the description of many of their topics.

Don't Know As predicted, the raters correctly identified the vast majority of items. Only 22 sentences remained unclassified with no particular domain attracting more *Don't Know* classifications. Only one sentence was rated as *Don't Know* by all three raters: *Many of those years were harsh and cruel.* Although from a History text, the sentence could equally well be attributed to Narrative given that the author seems to be voicing an opinion rather than an objective fact.

Number of Words Used

High inter-rater reliability is required to establish confidence that the number of words used by raters to assess the genre of sentences is suitably representative of experts' judgments. Following Hatch and Lazarton (1991), the adjusted correlation for three raters was $r = .660$, $p < .001$. For items for which *all three* raters correctly assessed the genre of the sentence, the correlation was $r = .732$, $p < .001$. The consistency across raters means that we can take the average number of words used by raters as the gold-standard representative of experts in assessments of the genre of sentences.

For the corpus as a whole ($N = 210$), the average number of words used by raters was 6.948 ($SD = 2.818$). As predicted, this is less than half the average length of sentences in the corpus. However, when we divide the corpus for the condition of *all raters giving correct judgments*/*other sentences*, the results show that significantly fewer words were required to *correctly* identify the genre (Correct: $N = 144$, $M = 6.419$, $SD = 2.407$; Incorrect: $N = 66$, $M = 8.101$, $SD = 3.256$; $F(1,208) = 31.140$, $p < .001$, $\eta^2 = .130$). This result suggests that a rater judgment of *fewer* than seven words is more likely to be correct, and a judgment of *greater* than seven words is more

Table 4: The three longest, misclassified sentences.

Domain	Classification	Sentence
Narrative	Don't Know	Friends in the barrio explained that the director was called a principal, and that it was a lady and not a man.
History	Narrative	The governor presided over an advisory council, usually appointed by the governor, and a local assembly elected by landowning white males.
History	Don't Know	We blow the whistle that's heard round the world, and all peoples stop to heed and welcome it.

likely to be *in*correct. The three sentences for which raters took the most words to arrive at the *wrong* domain are shown in Table 4.

To better understand the above result, we considered each domain individually. The results suggested that the seven-word average applied only to Narratives (Correct: N = 187, M = 6.808, SD = 3.029; Incorrect: N = 23, M = 9.870, SD = 4.808; F (1, 208) = 18.028, $p < .001$). There was no significant difference for correctly identifying domain using fewer words for the domains of History or Science. The similarity here between History and Science domains and the distinction from Narrative offers support to the conclusions of Graesser et al. (2002), McCarthy et al. (2007) and McDaniel et al. (1986). The result offers evidence that if an expert reader of a Narrative sentence has not become sufficiently aware of the sentence's domain by the seventh word that it is unlikely that subsequent words will make the reader any the more sure of the domain.

Discussion

In this study, we asked three experts in discourse processing to identify the genre of isolated sentences culled from a corpus of Narrative, History, and Science texts. Demonstrating high agreement, the raters in our study showed that expert readers could significantly identify the domain of over 80% of sentences. Further, our raters demonstrated that fewer than seven words (less than half the sentence) were required to correctly classify these sentences. Indeed, for the Narrative sentences, viewing more than seven words did not improve the accuracy of identifying the domain. These results suggest that the first half of sentences alone contains sufficient domain characteristics for skilled readers to begin the process of activating knowledge of text structure: a process which facilitates comprehension. Such research may lead to better understanding of how knowledge is represented and subsequently activated.

In addition, if only the first seven words of a sentence is sufficient for experts to recognize the text's domain, then computational approaches to text analyses may need to follow this lead. That is, text assessment for such features as readability, difficulty, genre, and cohesion may also need to be performed on just the first half of sentences because it is here that a significant part of human evaluation of the text seems to occur. More specifically, computationally evaluating an entire sentence may incorrectly assess the sentence's second-half as relevant to the reader's processing. In fact, this second half may be redundant or even noise in terms of reader activation of certain processing components.

Our results also showed that expert readers viewed many of the History and Science sentences as Narrative, suggesting that Expository texts tend to comprise a notable number of Narrative-like sentences. On the other hand, regardless of the domain from which sentences were taken, our raters were least likely to classify sentences as Science. This result sheds like on the heterogeneous compositionality of text, and provides significant implications for computational research. For instance, research in Text Mining, Genre Identification, and Information Retrieval tends to assume a high degree of homogeneity across texts and domains. Thus, computational approaches have tended to assume that the text as a whole is representative of the genre or text-type to which it has been assigned. The results of this study suggest that texts of any given domain may typically comprise sentences from many other domains. Understanding this diverse compositionality may lead to changes in how computational tools assess text searches and evaluations.

The compositionality of text is also a factor for research in reading development. Our results here suggest that for a text to be suitably representative of any given domain may require that the text contains a notable number of sentences more indicative of other domains. If a text does not contain this mixture of domain sentences, it is possible that a reader may have greater difficulty processing the text, as certain expectations may not be met.

In this study, we also addressed the question as to which genre best represents the History domain. Our results suggest that expert readers are as able to identify and distinguish History sentences as they are Science and Narrative sentences. This result supports the findings of Lightman et al. (2007), who found that History texts were distinct from both Science and Narrative texts. However, if we consider only the 39 misclassified sentences of the History domain, our results showed that 64% of these sentences were incorrectly assigned by our experts as Narratives, whereas only 18% of the sentences were identified as Science (and the remainder as *Don't Know*). Viewed this way, the result suggests that a notable portion

of History texts comprise Narrative like structures, a result that supports Duran et al. (2007), who found that History texts were more Narrative-like than Science-like. The categorization of History texts is important to cognitive science as a vast array of experiments typically assume that a History text is an Expository text. Consequently, experiments typically assume that History text will lead to similar results as Science text and different results from Narrative texts. The results of this experiment demonstrate that such an assumption could lead to erroneous analyses.

The results of our study lead us to two main areas of future research. First, a larger experiment is needed including participants of varying reading ability. Our results showed that skilled readers required fewer than seven words to successfully activate sufficient knowledge to recognize textual domains. Presumably, this activation skill is beneficial to reading and comprehension development. As such, we might expect that the number of words necessary to correctly recognize domains to be indicative of reading ability. Second, a detailed analysis of the form of the sentences at their point of recognition is needed. That is, we need to assess whether recognition stems mainly by way of the lexical items in the sentence, or by way of the structure of the sentence.

While much work remains to be done, our study demonstrates that genre recognition at the sub-sentential level is possible. Such recognition might provide a signature of reading ability, and as a consequence, a method of assessing reading ability. The major results of this study certainly provide sufficient initial evidence that such an approach is viable and that this paradigm can be further explored as an assessment of reading skill. Furthermore, there have been no previous investigations of how much text is required to recognize genre. This study indicates that very little text is actually required and that readers most likely activate information about text structure very early in the reading process.

Acknowledgements

This research was supported by the Institute for Education Sciences (IES; R305G020018, R305G040046). The authors would also like to acknowledge the contributions made to this paper by Art Graesser, Max Louwerse, Adam Renner, Zhiqiang Cai, Stephen Briner, Roger Taylor, Gwyneth Lewis, and Nick Benesh.

References

Bhatia, V. (1997). Applied genre analysis and ESP. In T. Miller (Ed.), *Functional approaches to written text: Classroom applications.* Washington, DC: USIA.

Clements, P. (1979). The effects of staging on recall from prose. In R.O. Freedle (Ed.) *New Directions in Discourse Processing* (pp. 297–330). Norwood, NJ: Ablex.

Duran, N.D., McCarthy, P.M., Graesser, A.C., & McNamara, D.S. (2007). Using Temporal Cohesion to Predict Temporal Coherence in Narrative and Expository Texts. *Behavioral Research and Methods*.

Graesser, A. C., Hauft-Smith, K., Cohen, A. D., & Pyles, L. D. (1980). Advanced outlines, familiarity, text genre, and retention of prose. *Journal of Experimental Education, 48*, 209-220.

Graesser, A. C., Hoffman, N. L., & Clark, L. F. (1980). Structural components of reading time. *Journal of Verbal Learning and Verbal Behavior, 19*, 131-151.

Graesser, A.C., Olde, B. A., & Klettke, B. (2002). How does the mind construct and represent stories? In M. Green, J. Strange, and T. Brock (Eds.), *Narrative Impact: Social and Cognitive Foundations.* Mahwah: NJ: Erlbaum.

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

Lightman, E.J., McCarthy, P.M., Dufty, D.F., & McNamara, D.S. (2007). *The Structural Organization of High School Educational Texts.* FLAIRS, 2007.

McCarthy, P.M., Renner, A.M., Duncan, M.G., Duran, N.D., Lightman, E.J., & McNamara. D.S., (in press). A Comparison of Two Models to Computationally Identify Topic Sentencehood. *Behaviorial Research and Methods.*

McCarthy, P.M., Graesser, A.C., & McNamara, D.S. (2006, July). *Distinguishing genre using Coh-Metrix indices of cohesion*. Paper presented at the Society for Text and Discourse conference, Minneapolis, MN

McDaniel, M. A., Einstein, G. O., Dunay, P. K., & Cobb, R. E. (1986). Encoding difficulty and memory: Toward a unifying theory. *Journal of Memory and Language, 25*, 645-656.

Meyer, B. J. F., & Wijekumar, K. (2007). Web-based tutoring of the structure strategy: Theoretical background, deisgn, and findings. In D. S. McNamara (Ed.), *Reading Comprehension Strategies*: *Theories, Interventions, and Technologies*. Erlbaum.

Oakhill, J., & Cain, K. (2007). Issues of causality in children's reading comprehension. In D. S. McNamara (Ed.), *Reading Comprehension Strategies*: *Theories, Interventions, and Technologies*. Erlbaum.

Otero, J., Leon, J. A., & Graesser, A. C. (Eds.), (2002). *The psychology of science text comprehension.* Mahwah, NJ: Erlbaum.

Tonjes, M.J., Ray, W., & Zintz, M.V. (1999). *Integrated Content Literacy.* New York: The McGraw-Hill Publishers.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Williams, P. J. (2007). Literacy in the Curriculum: Integrating Text Structure and Content Area Instruction. In D. S. McNamara (Ed.), *Reading Comprehension Strategies*: *Theories, Interventions, and Technologies*. Erlbaum.