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

Philip M. McCarthy (pmmccrth@memphis.edu)
Department of English, The University of Memphis, Memphis. TN 38152
Stephen W. Briner (sbriner@depaul.edu)
Department of Psychology, DePaul University, Chicago, IL 60614
John C. Myers (jcmyers@memphis.edu)
Arthur C. Graesser (a-graesser@memphis.edu)
Danielle S. McNamara (dsmcnamr@memphis.edu)
Department of Psychology, The University of Memphis, Memphis. TN 38152

Abstract

Genre identification is a critical facet of text comprehension, but very little is known about the process and information constraints of classifying texts by genres. In this study, higher-skill and lower-skill participants read 210 sentences from three genres. The words in the sentences were presented sequentially, one at a time. With each new word, participants decided whether the sentences came from a *narrative*, *science*, or *history* text. Both groups were able to correctly identify the genre by the third word of the sentence. Higher-skilled readers made their genre decisions more quickly and more accurately, and were also more precise in their selection of narrative texts. The study includes a computational model that uses text features from only the first three words of the sentences. The model reflects key features of the participants' genre classifications.

Keywords: genre recognition; reading skill; categorization

Introduction

Reading comprehension is greatly influenced by the genre of the text. Whether a text is a narrative, history, or science text influences the characteristics of the text, how the text is read, and can have a substantial influence on how well it will be understood (Bhatia, 1997; Graesser, Olde, & Klettke, 2002; Zwaan, 1993). More skilled readers utilize different strategies depending on the genre of the text (van Dijk & Kintsch, 1983; Zwaan, 1993) and training readers to recognize text structure helps to improve their comprehension (Meyer & Wijekumar, 2007; Oakhill & Cain, 2007; Williams, 2007). Once text genre is identified, it guides the reader's memory activations, expectations, inferences, depth of comprehension, evaluation of truth and relevance, pragmatic ground-rules, and other psychological mechanisms. For example, readers are more likely to scrutinize whether an event actually occurred in a history text, whereas that is not a particularly relevant consideration in most narrative fiction (Coleridge, 1985; Gerrig, 1993). In contrast, stylistic attributes are more important in literary narratives than expository texts (Zwaan, 1993).

Better understanding the nature of text genre and its effects on comprehension is important for theories of text comprehension as well as interventions to improve comprehension. If readers are indeed using different strategies to process different genres of text, then it is important to understand this process and potential information constraints during the course of genre identification.

We ask five questions in the current study. First, how quickly (in terms of number of words) do readers identify the genre of a text? Second, what types of errors (i.e., genre misclassifications) do readers make when identifying genres? Third, does the process of genre identification depend on reading skill? Fourth, what textual features (e.g., syntax, lexical choice) influence genre identification? And fifth, can a computational model categorize genre as humans do, using information available in only the initial words of sentences?

In McCarthy and McNamara (2007), we conducted a pilot study to provide a preliminary answer to our first two questions. Three experts (i.e. published authors) in the psychology of discourse processing were asked to identify the genre of isolated sentences culled from a corpus of narrative, history, and science texts. The experts had high inter-rater agreement (min = 90%) and required less than half the words in the sentence to accurately identify genres (accuracy as measured by F1, a standard index that considers both recall and precision: Narrative = .82; History = .84; Science = .82). The results further showed that these experts often classified many history and science sentences as narrative, suggesting that expository texts tend to be composed of a notable number of narrative-like sentences. On the other hand, science-like sentences were the least likely to be misclassified into other genres, suggesting the science-like sentences seldom occur in the non-science genres.

The current study builds on the study conducted by McCarthy and McNamara (2007) by including a larger sample of participants, an independent assessment of reading ability, a measure of *time on task*, and recording accuracy in terms of *number of words* used. We also construct a computational model based on our results. We use the model to investigate what information could be present in the initial words of sentences such that it can provide participants with sufficient information to make a genre evaluation. The question of whether or not we could build a computational model was important because if such relatively shallow features as syntax and word frequency information are sufficient for readers to distinguish genres, then it is reasonable to assume that (at some level) readers are using this information to process and categorize input. Our computational model sheds light on the features of the text that most likely influences readers' genre classifications.

Corpus

The corpus (as used in McCarthy & McNamara, 2007) comprises 210 sentences including 70 sentences from each genre (narrative, history, science). The sentences were compiled using corpora from two prior studies (Duran et al., 2007; McCarthy et al., 2007), which included 23 paragraphs each of 3, 4, and 5 sentences in length. Sentence selection from these paragraphs was guided by studies indicating that topic sentences are processed differently from other sentences in a paragraph (e.g., Kieras, 1978) and that topic sentences are more likely to occur in the paragraph initial position (Kieras, 1978, McCarthy et al., 2007). Thus, we sampled an equal number of paragraph-initial sentences and paragraph-non-initial sentences. For the latter, we used the third sentence of each paragraph because all paragraphs contained a third sentence and because third-sentences are presumably less closely related via co-reference and other semantic attributes to first-sentences than first-sentences are to second-sentences. To ensure that participants viewed sentences of approximately equal length, we included sentences that were within one standard deviation of the average length among 414 candidate sentences (M = 15.44 words; SD = 7.11). This resulted in a corpus of 210 sentences, equally representing the three genres and the initial/non-initial sentence dichotomy[1]. An example of a history sentence was *Because of the fragmented nature of Mayan society, the different cities frequently went to war*. An example of a narrative sentence was *The sweat from my other hand reduced the answers on my palm to a blue smudge*. An example of a science sentence was *Likewise, it's easier to express the concentration of a solution as the number of moles of material dissolved in it.*

Methods

Participants. There were 22 participants (Male = 10, Female = 12; M = 24.1 years old) who received $50 in exchange for participation. All participants were native English speakers. Fifteen participants were undergraduate students, five participants were graduate students, and two participants identified themselves as non-students.

Assessments. To assess reading skill, we used the Gates-MacGinitie (GM) reading test, a multiple-choice test consisting of 48 questions designed to measure reading comprehension. We used the level 10/12 version of the test, which has a reliability of .93 (MacGinitie et al, 2002).

Participants' genre recognition was evaluated using a Visual Basic program modified from McCarthy and McNamara (2007). The program consisted of three parts: *instructions*, *practice examples*, and *testing*. After viewing the instructions, participants were given six practice sentences. For the testing section, each participant evaluated 210 sentences presented in a random order. The program displayed the first word of the first sentence in a text window. Participants were required to select which genre they thought the sentence fragment belonged to. Participants made their selection by clicking on one of four on-screen buttons: *Narrative*, *History*, *Science*, and *Don't Know*. The buttons' position was randomized such that the genre choice could appear in any of the four buttons. Upon selecting one of the buttons, the mouse cursor returned to a central position so that each button was always equidistant from the start point of the cursor. As soon as a genre choice had been 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 …* the word *Yes* would appear as *Yes* + comma.) After 10 seconds, if the participant had not yet made a choice, a new word automatically appeared in the text window along with a message informing the participant of the new word. Three variables were recorded: *genre choice*, *accuracy*, and *time on task*. 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 had been presented. Thus, participants viewed a minimum of three words.

One difference between the current study and McCarthy and McNamara (2007) is the calculation of number of words used in the sentence by the participant to recognize genre. In the previous study, if *five* words were viewed then the number of words to make a decision was considered *five*. That is, the final two words, which are all evaluated as the same genre, were also included in the count. However, in this study, the number of words is calculated based on the first time the participant makes a choice that is repeated three times. Thus, in the above case, the number would be *three* words. Because the final two viewings merely repeat the participants' decision, we deemed it more accurate not to include the two extra viewings in the count of words needed to make a correct decision. As such, the McCarthy and McNamara (2007) report of *seven* words as being sufficient for experts to accurately classify genres could be viewed as *five* words according to the method employed in the current paper.

[1] We modified one science sentence that was a sentence fragment, changing *Taking no joy in life, looking forward to nothing, wanting to withdraw from people and activities* to <u>Examples are</u> *taking no joy in life, looking forward to nothing, wanting to withdraw from people and activities.*

Results

Subject Analysis

Our results showed that participants typically needed only a sentence's first three words to make their decision on genre (overall words used: M = 3.35, SD = 1.50; words used in correct assessments only: M = 3.33, SD = 1.45). The average accuracy of genre categorization was high (Recall: 0.86; Precision: 0.71; F1: 0.77), and this accuracy was consistent across the three genres (see Table 1). These results are consistent with our previous study (McCarthy & McNamara, 2007).

The magnitude of the correlation between *reading skill* (GM) and *words used* was moderate (r = .37, p = .09), as was the relationship between *words used* and *accuracy* (in terms of correlations with F1 participant evaluations, Science: r = .43, p < .05; Narrative: r = .37, p = < .09, History: r = .37, p < .09). We examined the results more closely by dividing the participants into two groups based on a mean split of the Gates-MacGinitie test scores (M = 24.00; SD = 9.14). Using these values, 13 participants were designated as lower-skill (LS) and 9 participants were designated as higher-skill (HS). Differences in Gates-MacGinitie test scores were analyzed using Levene's test for equality of error variances. No significant differences between groups were detected (p > 0.5), indicating that the groups are suitable for comparison.

Table 1: Accuracy of genre evaluation

Genre	Accuracy	Mean	SD
Narrative	Recall	0.86	0.09
	Precision	0.71	0.12
	F1	0.77	0.09
History	Recall	0.71	0.14
	Precision	0.76	0.09
	F1	0.72	0.11
Science	Recall	0.67	0.12
	Precision	0.88	0.09
	F1	0.75	0.11

We conducted an exploratory Analysis of Variance (ANOVA) to determine which of 22 variables best distinguished the reading skill groups. The analysis revealed that 7 variables significantly distinguished the two skill groups (p < .05) and 4 variables were marginally significant (p < .1). The most predictive variables were *narrative-precision* (Lower skill: M = .66, SD = .12; Higher Skill: M = .79, SD = .08; F = 7.55, p = .01, η^2 = .27) and *time for third word in history sentences* (Lower skill: M = 1.01, SD = .29; Higher Skill: M = .72, SD = .21; F = 6.72, p = .02, $P.\ \eta^2$ = .25).

The *narrative-precision* variable suggests that higher-skilled readers tend to be better at *not* classifying non-narrative sentences as narratives. In other words, skilled readers know better when a sentence is *not* a Narrative. These readers' greater accuracy may be because they are prepared to use more words than the Lower-skilled readers. However, a t-test revealed no significant differences between the number of words required by lower-skilled readers (M = 2.97; SD = 1.21) and higher-skilled readers (M = 3.85; SD = 1.68), t > 1.0, p > .1. Despite the lack of a significant difference between the higher-skilled and lower-skilled readers in terms of words used, the direction of the difference suggests that lower-skilled readers may too easily assume the direction or nature of the sentence discourse.

The variable, *time for the 3rd word in history sentences*, indicates the time on task for judging the third word of history sentences for correct decisions. Lower-skilled readers took significantly *more* time on this word. Indeed, *time on task* negatively correlated consistently with GM reading skill across all three genres for both 2nd words of sentences (Narrative: r = -.427, p = .05; History: r = -.443, p = .04; Science: r = -.523, p = .01) and 3rd words of sentences (Narrative: r = -.596, p < .01; History: r = -.606, p < .01; Science: r = -.500, p = .02). These results suggest that higher-skilled readers may be able to more quickly integrate new information.

Taken together, the results suggest that higher-skilled readers are more able to quickly and accurately process sentential information, using as few as the first three words. This advantage appears most evident in two features: on the 3rd word of sentences (all other word positions demonstrated weaker results); and in the accuracy of the precision of Narrative discourse. One further variable of interest is that higher-skilled readers may be prepared to use more words before making genre decisions. This final point is supported by our previous study (McCarthy & McNamara, 2007) in which expert readers (and therefore, presumably higher in ability than those who participated in this study) tended to use at least two more words than those who participated here. However, caution should be taken with these conclusions for two reasons. First, regarding word count, different experiments cannot easily be compared; and second, a step-wise multiple regression revealed that only the time on task for 3rd words of history sentences variable contributed to the model (adjusted R-square = .336). As such, the results of this and our previous study can only indicate the direction of subsequent experiments, which may shed more light on the relationship between reading skill and genre recognition.

Item Analysis

Of the 210 sentences in this study, only 4 (2%) failed to be correctly evaluated by any of the participants. For instance, the history sentence "*I had vainly flattered myself that without very much bloodshed it might be done*" was evaluated by all participants as a narrative; and the science sentence "*Hindi is the most widely used, but*

English is often spoken in government and business" was evaluated by 20 participants as history and by 2 as narrative. A further 33 sentences (16%) were correctly categorized by all the participants. For instance, the narrative sentence "*Why, I wouldn't have a child of mine, an impressionable little thing, live in such a room for worlds*" resulted in no misclassifications. For over half the sentences (55%) at least 19 of the 22 participants correctly evaluated the genre. For instance, the science sentence "*In areas with hard water, many consumers use appliances called water softeners to remove the metal ions"* recorded only three misclassifications. Conversely, only 10% of the sentences received less than 6 correct evaluations, an example being the narrative "*The Empress of Russia looked dressed for war, Igor thought.*"

The item analysis also showed that the sentences that received the highest accuracy in terms of categorization were likely to require fewer *words* for such categorization to be made. Thus, there was a negative correlation between the percentage of participants who correctly evaluated a sentence and the number of words needed to correctly categorize the sentence (r = -.639, p < .001). For example, "*Chemical weathering processes change the chemical composition of rocks"* was correctly identified as a science sentence by all of the participants and required an average of only 1.23 words to be identified. In contrast, "*However, this process was too slow to satisfy the Renaissance demand for knowledge and books"* was correctly categorized by only 27% (n = 6) of the participants and required 10 words to be correctly identified as a history sentence.

The results of the *time on task* demonstrated similar results. Specifically, there was a negative correlation between the percentage of participants who correctly assessed a sentence and average *time on task* for assessment (r = -320, p < .001). The results for both *words used* and *time on task* were consistent across the genres of Narrative (words: r = -.613, p < .001; time: r = -.466, p < .001); History (words: r = -.701, p < .001; time: r = -.404, p < .001); and Science (words: r = -.578, p < .001; time: r = -.257, p = .034).

Thus, consistent with the results reported by McCarthy and McNamara (2007), viewing more words does not lead to greater genre classification accuracy. This result indicates that if a sentence does not contain genre-specific features early in its structure, then it is also unlikely to contain those features later in its structure. The results for *time on task* indicate that sentences that are more accurately classified are also more quickly classified. We can presume that the quicker the decision, the less the processing necessary to make the correct decision. Thus, we did not observe a time/accuracy tradeoff.

Collectively, the results suggest that most sentences from the three genres can be accurately categorized in relatively few words and relatively little time. However, the variation within this accuracy suggests a continuum of *sentence-categorization difficulty*. That is, the first few words of sentences can often be sufficiently non-prototypical or ambiguous to reduce the likelihood of correct reader categorization. As such, it is feasible that the construction of the initial aspects of a sentence may significantly affect sentence processing, with less prototypical constructions causing readers to activate non-optimal schema.

Computational Model

The results of our previous experiment (McCarthy & McNamara, 2007) provided evidence that genre identification could be accomplished with less than half the words of sentences. However, given such little

Table 4: The 14 significant genre predictor variables with highest F-values for genres of Narrative, History, and Science

Dependent Variable	Narrative	History	Science	F	Effect Size
Past tense verbs incidence	177.3 (168.12)	68.18 (136.01)	13.33 (65.98)	20.01***	0.23
Ratio pronouns/noun phrases	184.04 (167.93)	51.14 (119.51)	38.33 (105.42)	17.29***	0.20
Syllables incidence	3.7 (0.86)	4.89 (1.54)	4.94 (1.46)	13.21***	0.16
Plural nouns incidence	21.28 (82.36)	34.09 (97.31)	113.33 (173.14)	7.62**	0.10
Singular proper nouns incidence	70.92 (169.34)	136.36 (209.66)	20.00 (79.97)	6.22**	0.08
Age of Acquisition (content)	42.26 (107.92)	60.85 (146.39)	146.06 (201.49)	5.96**	0.08
Mean Meaningfulness	362.35 (68.78)	305.48 (99.77)	312.42 (94.45)	5.72**	0.08
Nouns, singular/mass (incidence)	83.33 (144.34)	87.12 (144.85)	178.33 (214.29)	4.71*	0.06
Minimum Meaningfulness	182.32 (204.01)	150.27 (199.39)	70.28 (155.74)	4.65*	0.06
Verbs: non-3rd person incidence	14.18 (68.01)	15.15 (70.24)	66.67 (134.69)	4.59*	0.06
Mean Imageability	345.62 (68.56)	296.79 (92.52)	319.00 (79.45)	4.21*	0.06
Cardinal numbers incidence	14.18 (68.01)	53.03 (123.33)	6.67 (47.14)	4.00*	0.06
Verb, past participle incidence	0.00 (0.00)	0.00 (0.00)	26.67 (91.35)	3.87*	0.05
Verb phrases incidence	253.19 (199.32)	147.73 (187.52)	168.33 (196.69)	3.82*	0.05

Note: *** p < .001; ** p <.010; * p < .050; SD appear in parentheses; effect sizes calculated as η^2

discourse information, we examined whether a computational model based on only lexical and syntactic features (i.e., the information used largely by participants) provided similar results. If the model replicated the results found with humans, then it potentially provides evidence that participants use such sentential features when processing text. More specifically, if readers utilize shallow features to identify genre, this would suggest that readers may be making such categorization extremely early in the sentence. Identifying where and how participants are making their categorization thus informs theories of reading comprehension as well as research in reading strategies.

To address our computational question, we conducted a number of basic assessments, suitable for sentence level analysis, using the first three words of each sentence in the corpus. We selected the conservative size of the first three words of the sentences because this was the lowest average number of words for any of the groups: (i.e., the lower-skill group: M = 2.98 words, SD = 1.24). Our computational assessments included *word frequency values* (from the Celex data base), *word information values* (from the MRC data base), *parts of speech frequency counts* (based on the Charniak parser), and a *syllable count*.

To guard against issues of over-fitting and colinearity caused by applying multiple predictor variables, we followed established procedures of training and testing the algorithm (see Witten & Frank, 2005; McCarthy et al., 2007). The corpus was randomly divided into a training set (67%) and a test set (33%). Using the training set, we conducted an ANOVA to identify and retain only those variables that significantly distinguished the genre groups. We then conducted correlations among these variables and eliminated variables that presented problems of colinearity using $r \geq .7$; the variable with the higher univariate F-value was retained and the lower eliminated. Of the 16 remaining variables, the 14 with the highest univariate F-values were used in a discriminate analysis; there was an item to predictor ratio of 10:1 (see Table 4, above).

When the generated algorithm was applied to the *test set* data, the model significantly predicted accuracy (χ^2 = 22.48, $p < .001$). The model also predicted the data set as a whole χ^2 = 88.20, p < .001 (see Table 5). The results of the discriminant analysis suggest that the first three words of a sentence hold sufficient *syntactic* and *word level information* to significantly distinguish genre. The result is particularly impressive when considering the many constraints of the algorithm as opposed to those of the participants. First, the model is based on only the first three words of each of the sentences, whereas the accuracy of the participants includes the many instances where more than three words were used (indeed, the higher-skill group averaged closer to four words, 3.85, to correctly assess genre). Second, the algorithm included no predictors of world knowledge, which we can assume the participants would have. Thus, when participants see a number such as 1776 they are presumably more able to interpret this as an historical date. Third, even though word frequency was included as a predictor, the results are based on *frequencies in general* rather than genre specific. We can hypothesize that word information relevant to specific genres would enhance the accuracy of the prediction. For instance, we might assume that participants have knowledge that *cannon* is a word associated with history whereas *nucleus* is a word associated with science.

While we cannot claim that the model matches human performance, it is worth noting that the narrative precision evaluation for the test set (.74), all data (.73), and for participants (.71) are highly similar. Given that the human narrative precision variable correlated highly with reading skill (r = .520, p = .002), it is reasonable to assume that the computational model might reflect some aspects of reader strategy, at least in its propensity to correctly reject non-narrative decisions for narrative sentences. Additionally, the model's false alarms for narratives were similar to those decisions made by humans: that is, false alarms were less likely to be science decisions (History = 11; Science = 6).

Discussion

In this study, 22 participants identified the sentence genres of 210 sentences. The results indicated that both higher- and lower-skilled readers used about three words to accurately identify genres. Two primary variables related strongly to participants' reading ability: *Narrative-precision* and *Time on Task* for the 3rd word (i.e., typically the decision making word). Thus, higher-skilled readers are less likely to think a sentence is a narrative when it is not, and they also require less time to make their decisions.

Table 5: Recall, precision, and F1 results for computational model (test set; all data) compared to participant's performance

	Narrative			History			Science		
	Recall	Precision	F1	Recall	Precision	F1	Recall	Precision	F1
Test set	0.61	0.74	0.67	0.23	0.33	0.27	0.60	0.38	0.46
All data	0.67	0.73	0.70	0.46	0.52	0.49	0.71	0.59	0.65
Participants	0.85	0.71	0.77	0.71	0.76	0.72	0.68	0.87	0.76

The results of this study combined with our pilot study (McCarthy & McNamara, 2007) provide intriguing results. The results suggest 1) that a wide range of readers can accurately categorize genres at the sub-sentential level; 2) that as few as the first three words of a sentence may be all that is required for that assessment to occur; 3) that genre recognition may be indicative of reader ability; and 4) that features such as *time on task*, *accuracy*, and *word count* may be the indicators of reading ability. Further, the computational modeling results suggest that lexical and structural sentence features of just the first three words can also significantly classify sentences by genre. This result may help better identify prototypical and non-prototypical sentences in text. Such a model would also be useful for various computational procedures such as text mining, automatic summarization, and automatic web-genre classification.

The research presented here offers an interesting and promising direction toward a better understanding of how genre knowledge is represented and subsequently activated. We plan to use this knowledge to better establish our *genre identification paradigm* as an assessment of reading skill, and even as a possible intervention for reading development. This future research will need to consider such aspects as prior knowledge as well as consideration of which words in the sentences are selected for assessment; that is, do sentence *endings* have the same effect as sentence *beginnings*? While much remains to be done, the results presented here offer an exciting new perspective on the nature of text and the possibilities of reading skill assessment.

Acknowledgements

This research was supported in part by the Institute for Education Sciences (IES R305G020018-02) and in part by Counter-intelligence Field Activity (CIFA H9c104-07-C-0014). The views expressed in this paper do not necessarily reflect the views of the IES or CIFA.

References

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

Coleridge, S. T. (1985). *Biographia Literaria*: *Samuel Taylor Coleridge*, H. J. Jackson (Ed.), Oxford.

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. *Behavior, Research and Methods*. 39, 212-223.

Gerrig, R. (1993). *Experiencing narrative worlds: On the psychological activities of reading*. Cambridge, MA: MIT Press.

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.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2002). *Technical report for the fourth edition, Gates-MacGinitie reading tests*. Itasca, IL: Riverside.

McCarthy, P. M., & McNamara, D. S. (2007). Are seven words all we need? Recognizing genre at the sub-sentential level. In D. S. McNamara and G. Trafton (Eds.), *Proceedings of the 29th annual conference of the Cognitive Science Society* (pp. 1295-1300). Cognitive Science Society.

McCarthy, P. M., Renner, A. M., Duncan, M. G., Duran, N. D., Lightman, E. J., & McNamara. D. S. (in press). Identifying topic sentencehood. Behavior, Research and Methods.

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.

Meyer, B. J. F., & Wijekumar, K. (2007). Web-based tutoring of the structure strategy: Theoretical background, design, and findings. In D. S. McNamara (Ed.), *Reading Comprehsion Strategies*: *Theories, Interventions, and Technologies* (pp. 347-375). Erlbaum.

Oakhill, J., & Cain, K. (2007). Issues in causality in children's reading comprehension. In D. S. McNamara (Ed.), *Reading Comprehsion Strategies*: *Theories, Interventions, and Technologies* (pp.47-72). Erlbaum.

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

Williams, P. J. (2007). Literacy in the curriculum: Intergrating text structure and content area instruction. In D. S. McNamara (Ed.), *Reading Comprehension Strategies*: *Theories, Interventions, and Technologies* (pp. 199-220). Erlbaum

Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques.* San Francisco, CA: Morgan Kaufmann Publishers.

Zwaan, R. A. (1993). *Aspects of literary comprehension.* Amsterdam: John Benjamins.