Memory for Narrative and Expository Text: Independent Influences of Semantic Associations and Text Organization

Michael B. W. Wolfe
Grand Valley State University

The author examined memory for text in terms of the independent influences of semantic knowledge associations and text organization. Semantic associations were operationalized as the semantic relatedness between individual text concepts and the text as a whole and assessed with latent semantic analysis. The author assessed text organization by simulating comprehension with the construction integration model. Text organization consistently accounted for unique variance in recall. Semantic associations strongly predicted expository recall and predicted narrative recall significantly but to a lesser extent, even when the familiarity of the narrative content was manipulated. Results suggest that prior semantic associations and novel associations in the text structure influence memory independently, and that these influences can be affected by text genre.

When comprehending and recalling a text, readers rely on both the text itself and prior knowledge in a variety of ways. To gain a complete understanding of text comprehension and memory, researchers need to understand how these influences relate to one another. One aspect of prior knowledge that is important is the aspect of associations in semantic memory. Models of text processing indicate that semantic associations are important for activating relevant information that the reader may use to aid in comprehension (Kintsch, 1998; Myers & O'Brien, 1998). Semantic associations are likely to be important in predicting text memory as well. Another important factor in determining text memory is the organization of concepts in the text itself. Prior research has indicated that during comprehension, readers form a mental representation of text information in which the constituent elements (concepts) of the sentences are parsed into propositions, which indicate how the elements are related in the sentences. These propositions are then connected together across sentences to form a textbase representation (van Dijk & Kintsch, 1983). The textbase is a mental representation of a text that reflects the organization of the text elements as dictated by the text structure. In this article, semantic associations and text organization are evaluated in terms of their independent influence on memory for text. Recall of text information is predicted by computational models that assess semantic associations (Landauer & Dumais, 1997) and textbase strength (Kintsch, 1998). In addition, I examine these two influences on memory for both narrative and expository texts in order to assess the extent to which their relative influence may change under different processing circumstances.

To make concrete the factors of semantic associations and text organization, consider a text describing basic information about the human circulatory system that contains the following sentence: Blood leaves the heart through a big tube called the aorta, which is an inch thick. This sentence contains several text elements (concepts): LEAVES, BLOOD, HEART, TUBE, BIG, AORTA, INCH-THICK. Different theories of text processing provide means by which the relative strength of these elements in memory can be predicted. An analysis of the propositional structure of the sentence, for example, reveals the following complex proposition (van Dijk & Kintsch, 1983):

```
LEAVES
  \|-- BLOOD
    \  \-- HEART
        \  \-- BIG
            \  \-- TUBE
                \  \-- AORTA
                    \  \-- INCH-THICK
```

Note that the text elements in this proposition represent the content words from the sentence and refer to concepts (not literal words). Also, the elements are connected according to the text structure, which may or may not correspond to the relationships among these elements in prior knowledge. Memory for this sentence can be predicted on the basis of the extent to which text elements are central in terms of this propositional structure, in which case one would predict that leaves and tube would be well remembered. An analysis of the semantic associations involves considering the extent to which each of these text elements is important to the general topic of the circulatory system. By an intuitive judgment, heart, blood, and aorta seem to have the highest association with the general topic, and one would predict that they would be well recalled. They are related to the topic of the circulatory system, regardless of the role that they play in this particular sentence. Thus, using these semantic associations would lead these elements to be well recalled independent of their role in the structure of the sentence.
It is likely that the influences of semantic association and text organization on text memory will vary with different types of texts. Research on comprehension strategies indicates that readers can adjust the extent to which they focus on information within a text versus incorporating prior knowledge depending on the genre to which a text belongs (Linderholm & van den Broek, 2002; Zwaan, 1994). There are, of course, many different types of texts, which may trigger a variety of processing strategies. In this research, I used a representative set of narrative and expository texts. The labels narrative and expository encompass a range of specific text types that can have different, and sometimes overlapping, purposes. In these experiments, narrative texts are texts that describe a sequence of events that are related to each other temporally and causally (Brewer, 1980). The current narratives center around one or more protagonists who engage in a series of actions in order to satisfy a goal. The expository texts are designed to provide basic factual information about the structure, functioning, or sequence of events pertaining to topics that might typically be learned in school, such as the human circulatory system, volcanoes, or the U.S. Civil War.

Drawing firm conclusions regarding the narrative–expository distinction is difficult, however, because several variables are typically confounded between these genres. They may differ in terms of the way the texts are structured, the causal coherence of information, the presence of a protagonist, and the circumstances under which readers normally read these texts. Furthermore, there are different types of texts within the narrative and expository genres (Brewer, 1980; Goldman & Bisanz, 2002; Goldman & Varnhagen, 1986). Thus, these experiments should be viewed as a starting point in examining the independent influence of semantic associations and text organization as a function of text type. Other text types will likely induce different processing strategies.

Assessing Semantic Association and Textbase Influences on Recall

I obtained semantic association estimates using the latent semantic analysis (LSA) model of semantic knowledge (Landauer & Dumais, 1997). LSA is a computational model of semantic knowledge in which word meanings are acquired from large amounts (over 115,000 paragraphs) of naturalistic text. The texts that comprise the semantic space used here encompass a wide variety of topics, including novels, language arts texts, science, and social studies. Word meanings are discerned on the basis of their contextual usage and represented as vectors in a high-dimensional semantic space. One can also represent texts as vectors by taking the centroid, or vector average, of the individual word vectors for the words that make up the text. The similarity of these vectors to each other (computed as the cosine between them) provides a measure of the semantic relatedness between words or texts and each other.

To assess the role of semantic associations in memory for text, one assumes that the words in a text create a semantic context, and that this context can be represented computationally as the LSA centroid of the text. For narrative texts, the semantic context is created by all of the words used to describe the states, events, characters, and so forth that make up the story being told. For expository texts, the semantic context is the topic that is being explained, such as the U.S. government or the circulatory system. The LSA similarity between each text element and the entire text indicates the semantic relatedness to the context. In general, words such as topic-relevant nouns have a higher relatedness to the context than verbs or predicates that carry little semantic content. For example, Table 1 presents the LSA cosines between the sample sentence I presented in the introduction to this article and the elements of the sentence. Consistent with the intuitive judgments mentioned earlier, the elements that are directly related to the topic of the circulatory system have the highest cosines.

I obtained predictions based on text organization using Kintsch’s (1998) construction integration (CI) model of comprehension. In order to simulate construction of a textbase, the CI model proceeds in two phases. In the construction phase, complex propositions (van Dijk & Kintsch, 1983) representing relations among elements in a sentence are connected to one another on the basis of shared text elements (known as argument overlap). The construction phase creates an interconnected network in which text elements that are related to one another in the text structure are connected, and unrelated elements are not directly connected. In the integration phase, activation is spread through this network in a series of cycles that simulate the sequential nature of text processing. In each cycle, activation settles in richly connected elements and flows out of poorly connected elements. The results of each spreading activation cycle are then stored in a long-term memory matrix that contains elements with self-strength values and connection-strength values to other elements. After comprehension of the entire text has been simulated, the long-term memory matrix contains the product of comprehension; each text element has a long-term memory strength value and connections to other elements to which it was directly connected in the text structure (see Kintsch, 1998, for computational details). Table 1 also presents the CI long-term memory strength values for the example sentence discussed in the introduction to this article. It is important to note that the CI model also contains mechanisms for simulating the incorporation of prior knowledge and inference generation. In this context, however, only textbase predictions are desired, so I do not consider these aspects of the model. Thus, this research does not constitute a test of the complete CI model.

Experiment 1

In Experiment 1, I examined the influence of semantic knowledge associations and text organization on memory for narrative

Table 1

<table>
<thead>
<tr>
<th>Elements</th>
<th>CI</th>
<th>LSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEAVES</td>
<td>.64</td>
<td>.33</td>
</tr>
<tr>
<td>BLOOD</td>
<td>.14</td>
<td>.60</td>
</tr>
<tr>
<td>HEART</td>
<td>.14</td>
<td>.62</td>
</tr>
<tr>
<td>TUBE</td>
<td>1.00</td>
<td>.43</td>
</tr>
<tr>
<td>BIG</td>
<td>.21</td>
<td>.26</td>
</tr>
<tr>
<td>AORTA</td>
<td>.21</td>
<td>.68</td>
</tr>
<tr>
<td>INCH-THICK</td>
<td>.21</td>
<td>.36</td>
</tr>
</tbody>
</table>

Note. CI = construction integration; LSA = latent semantic analysis.
and expository texts. One goal was to examine the extent to which semantic associations and text organization provide unique predictions of recall. The other goal was to examine how these factors may differ between narrative and expository texts. To address these goals, I assessed CI and LSA predictions in terms of their ability to predict recall for each after factoring out the effects of the other.

Method

Subjects. In return for course credit, 144 subjects in an introductory psychology course at Grand Valley State University participated. Twenty-four subjects read and recalled each text.

Materials. I used 18 texts: 9 narrative and 9 expository. The narrative texts ranged from 26 to 29 sentences and had an average of 295 words. All of the narratives centered around one or more protagonists who have a goal and engage in actions related to achieving the goal. The texts all had an outcome-embedded structure (Goldman & Varnhagen, 1986) in which a superordinate goal is established near the beginning of the story. This goal leads to a sequence of episodes in which subgoals are established, and actions and outcomes of subgoals lead to one or more additional subgoals, until the superordinate goal is resolved at the end. For example, in “The Two Fools” (Magliano & Millis, 2003), the protagonist, Bastianello, wants to find two people who are more foolish than his fiancée’s family members if he is to become part of the family. This goal prompts him to travel from home, where he meets people and decides whether they are more foolish than his fiancée’s family members. After encountering two such people, he returns home. I collected the narrative texts from different sources (Boning, 1978; Magliano & Millis, 2003; Parker & Scannell, 1990) and modified them to fit into a common structure and length.

The nine expository texts all presented basic information about a topic and were 25 sentences long with an average of 290 words. I selected topics that would be somewhat familiar to undergraduates. I structured each expository text in the same manner. First, an introductory paragraph introduced the topic and stated that there were three components or factors that were important to the topic. For example, in “The Circulatory System,” the three components listed were the heart, the blood vessels, and the blood. The introductory paragraph was 4 sentences long in all texts. Next, three 6-sentence paragraphs described each of the three components or factors related to the topic. Finally, a 3-sentence closing paragraph reiterated the point that the three components or factors were important. In Goldman and Bisanz’s (2002) framework for classifying scientific texts, these texts fell specifically under the textbook genre, because they focus on well-established content and are designed to convey the content to the reader. I wrote all texts using various encyclopedias as references.

Procedure. All subjects participated in groups of 1–7. Text type (narrative vs. expository) was a between-subjects variable. Each subject read and recalled three of the nine texts of a particular type. Text order was counterbalanced such that each text appeared with each other text, and in each serial position, an equal number of times. On arrival, subjects received an experimental packet with instructions on the first page. The instructions were read aloud to subjects, and stated that subjects would read three short passages at their own pace, and that after that, they would perform some tasks related to the passages. Each text was printed on a separate sheet of paper. After finishing a text, subjects were not allowed to return to it. After finishing the third text, subjects turned the page, and a recall sheet instructed them to write down as much as they could remember about the first text. The title of the text was used as a recall cue. The recall task was self-paced. After completing recall of the first text, subjects turned the page and recalled the second text, and so on. After finishing recall of one text, subjects were not allowed to return to it. Approximately 10–30 min. separated the reading and recall phases for each text, depending on the serial position of the text. The entire session lasted approximately 20–45 min.

Results

Each text element of each passage was coded separately for gist recall. Gist was defined as recalling the exact word or phrase that represented the text element or a close synonym. For each text, 2 coders coded the recall protocols until agreement was at least 90%, at which point 1 coder coded the remainder of the subjects for that text. Disagreements were resolved through discussion. The average agreement between coders was 91%. Each text element that was recalled more than once was counted only once.

Model simulations. To generate LSA estimates of the semantic relatedness between text elements and the context created by the entire text, cosine data were generated from the LSA website (Laham, 1998). In all cases, the semantic space General-reading-up-to-first-year-college was used with 300 dimensions. For each text, the cosine between each text element and the entire text, including the title, was generated. For the CI simulations, complex propositions that represented the relationships between elements in each text sentence were generated. A connectivity matrix was established for each text in which complex propositions were connected on the basis of shared elements. Integration of the connectivity matrix into a textbase proceeded as specified by the CI model (Kintsch, 1998). The model parameters were constant across all simulations. The cycle size was set to seven elements, which indicates the minimum number of elements that participate in each cycle. After integration for each cycle, the single element with the highest activation in the cycle was carried over to the next cycle. The initial element- and link-strength values were set at one, which is the default setting for the CI program. Integration continued cycle by cycle until all complex propositions of the text were integrated. At this point, the long-term memory representation contained element- and link-strength values for all text elements in the text. The element and link strengths for the current texts were very highly correlated, and patterns of results were always the same, so only element-strength results are reported.

Free-recall data. In terms of overall amount of recall, subjects recalled more text elements from the narrative texts (M = 37.65, SD = 9.65) than they recalled from the expository texts (M = 23.69, SD = 6.00), F(1, 142) = 108.69, p < .0001. The primary analyses of concern, however, are the relative patterns of recall with respect to the predictions of the LSA and CI models. To examine recall patterns, I computed the average number of subjects recalling each element and assessed the ability of LSA and CI model predictions to account for these recall data using regression analyses. Partial correlations for LSA and CI indicated whether these predictors accounted for unique variability in recall of the text elements.

Partial correlations for LSA and CI are presented in Table 2, along with $R^2$ values, which indicate the total variance accounted for by both predictors. Several patterns in the data are notable. First, the CI strength values are predictive of recall data for both the narrative and expository texts. Second, the LSA predictions for the narrative elements are significant in four of the nine texts, suggesting that overall, LSA does a moderate job of predicting narrative recall. The LSA predictions for the expository elements are, however, highly predictive of recall. In addition to analyses of individual texts, the partial correlations from these regressions can be analyzed across texts within each text type. If the mean partial correlations across texts (reported in Table 2) are greater than zero,
this provides a further indication of the ability of LSA and CI to predict recall of these texts. For the narrative texts, both the LSA and CI partial correlations are significantly greater than zero, \(t(8) = 4.19, p = .003\), for LSA; \(t(8) = 12.00, p < .0001\), for CI. For the expository texts, both the LSA and CI partial correlations are also greater than zero, \(t(8) = 11.40, p < .0001\), for LSA; \(t(8) = 8.47, p < .0001\), for CI. Finally, one-way analyses of variance were used to assess the relative differences in partial correlations between narrative and expository texts. The LSA partial correlations are greater for the expository than for the narrative texts, \(F(1, 16) = 22.52, p < .0001\), whereas the CI partial correlations do not differ reliably between expository and narrative texts, \(F(1, 16) = 1.65, ns\).

Discussion

The results of Experiment 1 indicate that unique influences on memory for text are provided by semantic associations between text elements and the semantic context of a text and by the organization of text elements within a text. Predictions of text elements provided by the LSA and CI models accounted for unique variance in recall data for both narrative and expository texts. In addition, the LSA predictions of recall are much stronger for expository texts than for narrative texts. There are several differences between the texts that could potentially account for these results, however. One difference, which I explored in Experiment 2, is the fact that the narrative texts describe situations in which the reader does not know what events will take place. This is typical of narratives in general. Readers have world knowledge about objects, locations, story schema, and so forth, but they do not know a priori what events will transpire and how events will relate to each other. Expository texts, however, describe content in which the objects and relationships among objects are specifically familiar to readers, at least to the extent that they have some knowledge about the topic being discussed. For example, relations among components in the circulatory system will not change from one text to another. This difference can be characterized in terms of readers having topic-relevant knowledge for expository texts and no topic-relevant knowledge for narratives. It is possible that the topic-relevant knowledge in the expository texts triggers processing strategies that involve consideration of the general semantic relatedness of text elements to the topic.

Experiment 2

My goal in Experiment 2 was to examine the extent to which topic-relevant knowledge influences reliance on semantic associations in memory for text. Two complementary pairs of narrative texts were created that had the same text organization but for which the possibility that readers would have topic-relevant knowledge would differ. The text organization of the pairs of narratives was equated to ensure that the differences in the textbase could not account for differences in the memory results. Within each pair of narratives, the topic-relevant knowledge text contained content with predetermined relationships among the text elements, as is typically the case with educational or school-based texts, as well as with the expository texts in Experiment 1. One of these texts contained information about the circulatory system, and the other contained information about the U. S. Civil War. The complementary narrative within each pair (the non topic-relevant knowledge text) contained content in which the relationships among elements (the events) would not be specifically familiar to readers. If reading a text with topic-relevant content triggers readers to use information about semantic associations, then LSA should be more predictive of recall for the topic-relevant knowledge texts than for the non topic-relevant knowledge texts.

Method

Subjects. In return for course credit, 64 subjects in an introductory psychology course at David Lipscomb University participated. Thirty-two subjects read and recalled each text.

Materials. I created two complementary pairs of narrative texts for this experiment. Each text in a pair was structurally identical to the other text in the pair so that the texts would differ only as a function of content. As a result, the two texts in a pair had an identical number of sentences and elements and an identical propositional structure. For one pair of texts, the topic-relevant knowledge text, “John Wilkes Booth,” presented events leading up to and including the assassination of Abraham Lincoln. The motivation for Booth to assassinate Lincoln was established, and the events related to the assassination were described. This text also contained factual information relating to the U. S. Civil War. The non topic-relevant knowledge text, “Stuart Sayre,” was about a student athlete at a university who was upset with the school’s athletic director over the issue of athletic scholarships. At the end of the story, Sayre assaults the athletic director. Both texts were 23 sentences long. For the other pair of texts, the topic-relevant knowledge text, “Alex’s Adventure,” was a story about a man who builds a machine to shrink himself and gets sucked into a woman’s body. As he attempts to find his way out of the woman’s body, factual information about the heart and circulatory system is revealed. The non topic-relevant knowledge text, “Gary’s Adventure,” was a story about a boy who obtains his driver’s license and gets lost in a fictional city. As he attempts
to find his way back to the freeway, information about street names and landmarks in the fictional city are revealed. Both texts were 28 sentences long.

Procedure. Subjects were run together in a large classroom setting. Each subject read and recalled one of the two texts from each text pair. Texts were counterbalanced so that each text appeared either first or second equally often and with each of the texts in the other pair an equal number of times. Subjects received an experimental packet with instructions on the first page that were read aloud by the experimenter. The instructions stated that subjects were to read two short passages, and that after reading each passage they would perform another task for 30 s and would then write down the passage in as much detail as possible. Each passage was read once at the subject’s own pace, and the passage could not be returned to once the subject was finished reading. After reading, subjects turned to the next page and performed arithmetic problems for approximately 30 s. Then subjects turned the page again and recalled the passage in as much detail as possible. Recall instructions were the same as in Experiment 1. On completion of recall for the first passage, subjects turned the page again and read and recalled the second passage. The session lasted 35 min.

Results

Coding of recall data and simulations proceeded as in Experiment 1. Note that for each text pair in this experiment (Booth & Sayre, and Gary & Alex), the CI simulation values are identical for the two texts. In terms of overall recall, subjects recalled more text elements from the Booth text \((M = 44.5, SD = 10.92)\) than they did from the Sayre text \((M = 36.8, SD = 13.73)\), \(F(1, 62) = 6.25, p = .02\). There was no significant difference in the overall amount of recall for the Alex text \((M = 31.6, SD = 11.99)\) compared with the Gary text \((M = 27.1, SD = 9.45)\), \(F(1, 62) = 2.78, ns\). As in Experiment 1, I am primarily concerned with the relative patterns of recall in multiple regressions that predict recall using CI and LSA assessments of text organization and semantic associations, respectively. Partial correlation results are presented in Table 3. The results generally replicate the narrative results from Experiment 1. For the non-topic-relevant knowledge texts, the LSA predictions of recall are significant only for the Gary text, whereas the CI predictions are significant for both texts. In both of the topic-relevant knowledge texts, LSA predictions are low and non-significant, whereas the CI predictions are again significant in both cases.

Discussion

In pairs of narratives that were identical in terms of text organization, recall was predicted by the CI model regardless of whether the readers possessed topic-relevant knowledge. Recall was not predicted by assessments of the semantic relatedness between text elements and the context of the text (LSA). These results suggest that the differences in reliance on semantic relatedness found in Experiment 1 are not caused by differences in the content alone. Merely reading a text that contains content with predetermined relationships does not appear to trigger readers to consider the extent to which text concepts are related to the content domain being discussed.

General Discussion

These experiments demonstrate that the organization of text elements (concepts) in a text and the general semantic association of elements to the text context independently influence memory for text. Text elements were recalled better when they were more central in the text organization (as assessed by the CI model of Kintsch, 1998) and when they were more semantically related to the topic, or context, of the text (as assessed by the LSA model of Landauer & Dumais, 1997). In addition, in Experiment 1, semantic association effects differed as a function of text genre, with LSA predicting recall much better for expository texts than for narrative texts. These results suggest that semantic associations are important to text recall in general, but also that the genre of a text can trigger processing strategies that vary the extent to which these associations may be relevant to the task of processing and recalling a text. For expository texts, the prior semantic associations among text elements, which represented knowledge that the reader already possessed, play a critical role in determining text memory. For narrative texts, the novel associations established by the organization of concepts in the current story appear to be relatively more important than the general semantic associations. Finally, Experiment 2 demonstrated that specific topic-relevant knowledge, in the form of topics in which the relationships among elements are largely fixed (as in the expository texts in Experiment 1) is not sufficient to trigger reliance on semantic associations during recall. In pairs of narratives in which topic-relevant knowledge was manipulated, recall patterns mirrored the narrative results of Experiment 1, with CI predictions accounting for recall data much better than LSA predictions.

I chose the texts in these experiments to represent common types of narrative and expository texts. As such, they serve as a starting point for examining the influence of text organization and semantic associations on recall. These results should be viewed as somewhat exploratory, however, for two reasons. First, there are several differences between the texts used here (particularly in Experiment 1). Because I used representative narrative and expository texts, they differed in terms of organization, causal coherence, the presence of human characters, and vocabulary. Second, the distinction between narrative and expository texts can be difficult to determine in many instances, such as informative essays that have a temporal and causal structure to them. Also, readers are able to alter their processing strategies on the basis of specific reading purposes (Linderholm & van den Broek, 2002; Zwaan, 1994), so the patterns of results reported here may change within a specific text if the reading purpose changes. As a result, both text type and reading purpose will be important factors to examine in future studies.

<table>
<thead>
<tr>
<th>Text type</th>
<th>CI</th>
<th>LSA</th>
<th>(R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No topic-relevant knowledge</td>
<td>.28**</td>
<td>.04</td>
<td>.08*</td>
</tr>
<tr>
<td>Sayre (football)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gary (traffic)</td>
<td>.37***</td>
<td>.29**</td>
<td>.23***</td>
</tr>
<tr>
<td>Topic-relevant knowledge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Booth</td>
<td>.29**</td>
<td>.12</td>
<td>.10**</td>
</tr>
<tr>
<td>Alex (heart)</td>
<td>.33**</td>
<td>.00</td>
<td>.11**</td>
</tr>
</tbody>
</table>

Note. CI = construction integration; LSA = latent semantic analysis. * \(p < .05\). ** \(p < .01\). *** \(p < .001\).
References


Received June 2, 2004
Revision received September 3, 2004
Accepted September 20, 2004

ORDER FORM

Start my 2005 subscription to Journal of Experimental Psychology: Learning, Memory, and Cognition! ISSN: 0278-7393

$140.00, APA Member/Affiliate

$287.00, Individual Nonmember

$700.00, Institution

In DC add 5.75% / In MD add 5% sales tax

TOTAL AMOUNT ENCLOSED $ 

Subscription orders must be prepaid. (Subscriptions are on a calendar year basis only.) Allow 4-6 weeks for delivery of the first issue. Call for international subscription rates.

SEND THIS ORDER FORM TO:
American Psychological Association
Subscriptions
750 First Street, NE
Washington, DC 20002-4242

Or call (800) 374-2721, fax (202) 336-5568.

TDD/TTY (202) 336-6123.

For subscription information, e-mail:
subscriptions@apa.org

Send me a FREE Sample Issue

Check enclosed (make payable to APA)

Charge my: VISA MasterCard American Express

Cardholder Name

Card No. Exp. Date

Signature (Required for Charge)

BILLING ADDRESS:
Street

City State Zip

Daytime Phone

E-mail

SHIP TO:
Name

Address

City State Zip

APA Member # 

XLMA15