中 華 心 理 學 刊
民96，49卷，3期，245-269

Chinese Journal of Psychology
2007, Vol. 49, No. 3, 245-269

Memory for Narrative Texts: How do Parts of the Landscape Model Work

Yuhtsuen Tzeng

Center for Teacher Education & Institute of Curriculum, National Chung Cheng University

MS No.: 061106; Received: November 6, 2006; Revised: June 21, 2007; Accepted: June 30, 2007
Correspondence Author: Yuhtsuen Tzeng, Center for Teacher Education & Institute of Curriculum, National Chung Cheng University, 168, University Rd., Ming-Hsiung, Chia-Yi, 621, Taiwan（E-mail: ttcytt@ccu.edu.tw）

A computational approach is desirable for dealing with the complex nature of processes and memory representations in text comprehension. The study focused on a particular model of comprehension-the Landscape model-seeking to establish its validity. Study 1-a explored several conditions of the model and examined the model's ability to account for human recall data for short narratives. Parameter values were set for optimal conditions and cross-validated in Study 1-b for new story materials. The model showed high stability in accounting for these new texts, indicating its predictive power does not come from choices of particular parameter values. Several simulations in Study 2 revealed that the model's assumptions concerning the necessities of anaphoric and causal inferences were validated. The model's scheme of assigning activation values for different sources of activation was also supported because modifications of the current scheme reduced performance of the model. Limitations and several lines of future research are discussed.

Keywords: ***comprehension modeling, Landscape model, text memory***

The product of language comprehension is a memory structure with different levels of representation (Fletcher, 1994; Graesser, Olde, & Klettke, 2005; van Dijk & Kintsch, 1983). This representation serves many purposes for readers. One can use this representation to recall facts, to answer questions, to tell a story or even to assemble a piece of furniture. One major task of comprehension researchers is to figure out the nature of this representation and how readers arrive it. In order to build a coherent representation, readers constantly need to make inferences to fill in the gaps in a text. There are many possible types of infer-

Acknowledgements

The author is grateful for constructive comments from Dr. Ching-Fan Sheu and two anonymous reviewers. This project is partially supported by grants from National Science Council of Taiwan (NSC-92-2413-H-194-022). Readers who are interested in the Windows version of LS program can download it at the following website for free: http://www.education.umn.edu/EdPsych/Projects/LandscapeModel/default.html.

ence that a reader can make. We will focus on two particular types, namely anaphoric and causal inferences in accordance with the scope of this paper.

There is abundant evidence that readers routinely establish coherence by resolving anaphoric references during reading. Reading-time measures indicate that readers slow down when they encounter an anaphor that needs to be resolved (Garrod & Sanford, 1990; Haviland & Clark, 1974). Other studies have shown that anaphors may differ in their effectiveness in modulating activation of text elements. The more specific the anaphor is with respect to the identity of its referent, the faster readers will resolve it (Dell, McKoon, & Ratcliff, 1983; Gernbacher, 1989, 1990; Gernbacher & Faust, 1991). An example of a specific anaphor is direct repetition of an antecedent (e.g., burglar). An example of a less specific anaphor is a synonym (e.g., criminal). Several researchers also discovered that the distance between an anaphor and the referenced antecedent influences the speed of accessing the antecedent (e.g., Duffy & Rayner, 1990; Ehrlich & Rayner, 1983); antecedents that are more distant in texts take readers longer to resolve. It is clear that readers routinely make anaphoric inferences during reading and they do so in order to maintain the coherence of the text, which is essential for an integrated situation model.

Anaphoric relations may play a central role during the initial construction of coherence but other types of relations, such as causal ones, may add significantly (Kintsch, 1988; Trabasso, van den Broek, & Suh, 1989; van den Broek, 1994). Causal relations can explain causes and consequences of events or mental states and thereby give readers a sense of understanding which can not be achieved by co-reference (Fletcher, 1994). Readers may generate two types of causal inferences during reading. They may be backward, connecting the focal event to its antecedent, or they may be forward, anticipating future consequences of the focal event. Forward inferences are general considered to be highly elaborative therefore they are not necessary for maintaining the coherence of the text. As a result, they are not always reliably made during reading (Singer, 1994; van den Broek, 1990, 1994). In contrast, backward causal connections provide coherence between the focal statement and antecedent thus they are required for comprehension.

A model of backward inferences has been proposed by van den Broek (1994) emphasizing that readers are driven to establish a sufficient causal explanation for the focal event. Several lines of evidence have supported this hypothesis. The results of reading-time studies indicate that reading slows down when the reader encounters a statement for which local causal coherence is missing (Bloom, Fletcher, van den Broek, Reitz, & Shapiro, 1990; Klin & Myers, 1993; Thurlow, 1991). A similar pattern has been found when an explanation is provided but is not sufficient, for example when the current event requires the reinstatement of a goal or of other prior text in order to be sufficiently explained (Bloom, et. al., 1990; Dopkins, Klin, & Myers, 1993; Thurlow, 1991). Naming and lexical decision studies show that events that are predicted to be reinstated in order to provide sufficient explanation for a focal statement are more highly activated after the focal statement than before (Dopkins, et. al., 1993; Klin & Myers, 1993; Thurlow, 1991). Furthermore, such reinstatements occurred only when the immediately preceding statement did not provide full sufficiency (Thurlow, 1991). Similarly, verification times for facts from background knowledge are faster immediately following statements for which such background information provides sufficient explanation than following control statements (Singer, Halldorson, Lear, & Andrusiak, 1992).

Findings such as these suggest that readers slow down when they encounter situations in which the current event is not sufficiently explained by the immediately preceding statement, and that they reinstate or draw on background knowledge to activate information that provides the missing explanation. In addition to being generated during reading, these inferences appear to be incorporated into the memory representation of the text. Events that are hypothesized to have been reinstated during initial reading speed up recognition of their consequences in subsequent memory tasks (van den Broek & Lorch, 1993).

In sum, backward causal inferences are generated routinely during reading. As readers proceed through the text, they attempt to maintain causal coherence and sufficient explanation by making inferences. Findings such as these indicate that readers frequently engage in inferential processes to attain adequate coherence. Thus the presence or absence of anaphoric clarity and causally sufficient explanation influence both when different types of inferences will occur and what the functional contents of the resulting inference is likely to be.

Other researchers have questioned whether readers reliably make forward causal connections online despite that they might agree on making backward causal inferences (e.g., McKoon & Ratcliff, 1992). However, considering the empirical evidence described above and the fact that causality tends to capture the theme or macrostructure of narratives hence the essence of a situation model, it is reasonable to include causal inferences in descriptions of the online process for building situation models.

As more has been discovered about the cognitive processes and mental representations of reading comprehension, new interest has arisen for developing more precise models to explain how it works. A computational model serves this purpose well. There are several advantages of a computational model of comprehension. First, a computational approach is precise and logical. If the model is not completely specified or not consistent enough then it will not run. Second, since it requires detailed specifications, it forces researchers to be very explicit about their assumptions. This, in turn, helps them clarify ambiguity. Third, since every component of the model must be well articulated, one can easily test assumptions, eliminate unnecessary ones or falsify the theory. Fourth, the complexity of online inference and offline representation makes a computational approach the most plausible tool to keep track of all the components within the model and examine their interactions. Similar arguments have been articulated by Goldman, Golden, and van den Broek (in press).

In this research, a particular computational model of comprehension, the Landscape model, will be examined. The Landscape model (LS) is a connectionist model of comprehension with various new computational features that enable it to model many comprehension phenomena that earlier models could not fully handle (van den Broek, Risden, Fletcher, & Thurlow, 1996; van den Broek, Young, Tzeng, & Linderholm, 1999). One major goal of LS is to build a computational framework with psychological plausibility that can capture reader' mental representations resulting from text comprehension.

The LS has two memory stores: a working memory and a long-term episodic memory. For each processing cycle, the model takes one parsed text statement as input. Normally this text statement corresponds to a clause in a sentence with a main verb and can be depicted as a larger unit of propositions than the smallest atomic proposition. Once elements in a text statement are read, they are assumed to be active in working memory. Text elements are activated to different degrees. Explicitly read text elements are most active. Text elements also can be active as a result of inferential processes. LS assumes that elements are activated whenever referential or causal inferences are needed. This assumption is based on empirical evidence showing that these two are reliable online inferences and they play an important role in the construction of situation models. Free associated elements can be also activated if they are strongly related to text elements that are pertinent to research purposes.

In the LS model, whenever an element is activated during reading all the other elements currently activated are associated with it. This associative property produces two consequences. First, each element connects with a group of related elements thus forming a cohort. When one element is activated so are, to some extent, all the elements that have become associated with it. This feature is termed cohort activation. There is a *cohort parameter* in the model to determine the amount of cohort activation among the interconnected reading elements. Second, this associative feature also serves as a fundamental mechanism of relating online activation and offline representation. When two elements coactivate in working memory, a connection is built and this connection is stored in episodic memory. LS treats this building up of episodic memory connections as a learning phenomenon.

A learning algorithm such as the delta rule (McClelland & Rumelhart, 1985) is used to capture the online to offline relations. The delta rule is essentially a connectionist implementation of the Rescorla-Wagner model (Anderson, 1995), which is a theory of learning accounting for a wide range of data. Numerous studies have shown that the Rescorla-Wagner model predicts well learning from experience, and it has been successfully applied to a wide range of domains from behavior specifications to connectionist modeling (Gluck & Bower, 1988; for a review see Miller, Barnet, & Grahame, 1995; Wang, Johnson, & Zhang, 1998; Zhang, Johnson, & Wang, 1996). Although it has been shown that the Rescorla-Wagner rule has its limitations, such as its inability to handle complex non-linear problems (see Barto, 1990), it has been chosen because it is a directional learning rule, meaning that a connection from A to B is not the same as one from B to A. It is appropriate for modeling causal relations in text comprehension because causal relations are asymmetric. However, this model does not have a hidden layer and is not driven by any errors as in other typical connectionist models of cognition using the delta learning rule.

The model continuously builds a memory representation by forming associations between different elements as well as auto-associations of each element to itself. There is a *learning rate parameter* for updating the amount of new learning in episodic memory in the model. The LS model further assumes that connections between two elements will reach an asymptote as the number of their co-occurrence increases. The result of LS reading is a gradual emergence of a network representation with nodes and connections of various strengths. Moreover, these elements compete with each other for the privilege of predicting other elements and hence create expectancy. The amount of expectancy is related to the learning rate. The better a reading element can already expect/predict another element, the less association strength will be increased between them.

In addition to comprehension processes, LS model also contains a retrieval component. This retrieval mechanism follows a 2-step procedure. It retrieves the node with the highest total activation value across all reading cycles. Then the model searches all the connected concepts and retrieves the one with strongest connection with the retrieved node. This process is repeated until the model reaches a dead end. A dead end is defined as connection strengths between nodes below a certain pre-determined threshold. Ties are resolved by retrieving all the tied nodes in the sequence of their input order.

The LS model has been successful in accounting several comprehension phenomena. There is high correlation between the model's predicted activation values and readers' rating of activation (van den Broek et al., 1999). It can also predict recall by human readers (Kendeou & van den Broek, 2005). Variations of the LS model are capable of modeling human's detection of inconsistencies in a text (Linderholm, Virtue, Tzeng, & van den Broek, 2004). Furthermore, it has been shown to be able to mimic the important role of readers'prior knowledge (Kendeou & van den Broek, 2005). LS is a full-fledged computational tool- the LS program has been developed to provide a friendly interface for modeling research and it also serves as a platform for implementing and testing models and hypothesis (Tzeng, van den Broek, Kendeou, & Lee, 2005). Little is known about what makes the model work despite much evidence demonstrating its overall psychological validity. In this research, we collect additional evidence to support the LS model. More importantly, we take one step further to examine the internal mechanisms of the model and see how they work together to produce its predictive power.

Computational models of comprehension intend to describe and predict cognitive processes and the resulting representation of comprehension in a precise manner. In addition to their exact formulations, they are usually coherent enough to make sense of all the theoretical constructs in the presence of others. Precision and coherence, however, do not guarantee validity. A good computational model of comprehension should be able to account for human performance. Two major goals of this research were (1) to test validity of the LS model by comparing its performance to human recall data and (2) to examine the model's internal mechanisms. A free recall measure was chosen because it is has been under a long history of scientific scrutiny and it is also a frequent task in daily life. A more specific purpose was to explore the necessity of different components of the model such as different free parameters and default assumptions. Studies like these will advance our understanding of what is working and what is not working within the model and provide useful feedback for model building.

Study 1

The purpose of this study is to examine the behaviors of LS under various values of two free parameters: the cohort parameter and the learning rate parameter. There are no a priori constraints on what the proper values are for these parameters in a comprehension model. It is hence crucial to explore what their impacts are on the performance of the model in fitting human recall data. A bottom up approach was adopted to pick a set of best-fit parameter values for LS for two narrative texts in Study 1-a. The same set of parameter values then were cross-validated on two new narrative texts in Study 1-b. The results would constitute a test of the model's construct validity, namely, whether the model behaves in accordance with human data. The model will be judged as valid if its performance is affected by parameter values and it can predict human data at least within a certain range of its parameter values. If the model's behaviors are not affected by parameter values, there are no logical reasons to include these parameters and the model is considered to lack construct validity. Similarly, if the model does not predict human data, it is not doing what it is supposed to do. The model may not be able to predict human data at all parameter values because some parameter values might not be psychologically plausible. Nevertheless, the model has to do so for at least some parameter values.

Study 1-a

The purpose of this study was to explore a range of parameter values which would allow the model to fit human recall data.

Method

Materials. Four narrative texts from Trabasso, Secco, and van den Broek (1984) were used. They were short stories adapted from folktales that have been used in many comprehension studies. They are listed in Appendix 1. The short titles, numbers of sentences, and number of words are listed in Table 1 along with mean recall proportion, standard deviation, and recall ratio per story in this study. *The mean recall proportion* for each node in a story is calculated by counting the number of participants who recall a node divided by the total number of participants. *Recall ratio per story* is derived by computing the number of nodes that are recalled by at least one participant divided by the total number of nodes in a story. This measure treats all recalled nodes equally regardless of how many participants recall them. It was meant to mimic the current low retrieval threshold in LS which was set to zero to maximize model recall. The mean recall proportion for each story was not high (ranging from .24 to .53). However, the recall ratio per story was quite high (ranging from .92 to 1.0), reaching a ceiling effect.

These four stories (two for study 1-a and 1-b respectively) were parsed into concepts, which consisted of main verbs, protagonists, nouns and adjectives (for more details of the parsing principle see van den Broek et al., 1996). Referential connections were identified whenever there was argument overlap. Causal connections were identified following the procedures suggested by Trabasso et al. (1989) and van den Broek (1990). Input matrices for LS were prepared by combining explicit text inputs, referential and causal connections as sources of activation for each reading cycle using a default activation value scheme. That is to assign an activation value of five for explicit text input (mention), four for referential and causal connections, and three for enabling. The Tiger story, its parsing and its input matrix are listed in Appendix 2 as an example. These input matrices were submitted to an earlier version of the Landscape program in Mathematica 3.0.

Participants. Thirty-two college students were recruited from an Introductory Psychology class at a large Midwest US campus. They received course credit for participating. All participants were native speakers of English with normal (or corrected to normal) vision and without reading disability. The whole procedure took about 30 minutes.

Procedure. All stories were printed on a booklet with one story per page. Participants were told that this was a study of story comprehension. They were instructed to read all stories at their normal reading speed. Once they felt they understood the first story, they continued to the next. Participants were informed that there would be a recall task later. The order of stories each participant received was counterbalanced. During recall, participants were encouraged to recall as much as they could remember from the stories. A separate booklet with blank sheets was provided. Each sheet contained the titles of the stories in the same order as they were read. Once the participants finished reading all the stories, they wrote down all that they could remember on each sheet under the correct title. The whole session was conducted in one group. There were no time limits. Recall protocols were scored according to a gist criterion, i.e., participants would earn credits for recalling words or phrases that shared similar meanings of target text elements.

Six levels of the cohort parameter (0, .03, .06, .09, .12, .15) and six levels of the competitive learning rate parameter (0, .3, .6, .9, 1.2, 1.5) were selected to produce thirty-six conditions. These ranges of parameter values were far from comprehensive but they covered a substantial range of magnitudes and allow exploration of a parameter space likely to identify reasonable and stable parameter values. The model was run under these parameter conditions for each story, thus producing a total of seventy-two simulations for this study.

Two dependent measures were used to evaluate the performance of LS. The first dependent measure was the proportion of human recall predicted by LS. This measure was defined as the amount of LS recall divided by the amount of human recall. The amount of human recall was defined as the number of reading elements that are recalled by at least one participant. The amount of recall by LS was the number of reading elements retrieved by LS. The default state of LS was set to zero (i.e., any node with a positive connection with any previously retrieved node is retrieved). In some cases, LS recalled more elements than humans did. An adjustment was made to penalize the model's over-recall. The adjustment follows several steps. First, regular recall proportion was calculated by dividing the amount of LS's recall by human recall applying to recall elements that were part of human recall protocol. This is the hit rate. Second, the number of over-recalled elements was divided by the total number of human recall to produce a penalty measure for LS. This is the false alarm rate[1]. Finally, this penalty term (false alarm rate) was subtracted from the regular recall proportion to create an adjusted measure of recall propor-

1 I thank one of the reviewers for linking these measures to hit and false alarm rates.

tion. All recall proportions reported throughout this paper are adjusted measures. This penalty for the model's over-recall is necessary to bring the model to be constrained by human data and to prevent the model from taking advantages of large cohort values because the model tends to recall more with large cohort values. The range of recall proportion is always between 0 and 1. This measure is designed to tap how much information the model can retrieve from its memory representation.

The second dependent measure of the model's performance was the correlation of LS's long-term memory strength and human recall frequency. The rationale was that the more people recall a particular node, the stronger its memory trace must be. Pearson correlations between the model's long-term memory strengths and human recall proportions (the number of participants who recalled a node over the number of total participants). This is a measure to evaluate LS's ability in capturing the transformation from online activation into a stable memory representation. It does not rely on current retrieval mechanisms in the model. It is based on the memory representation before the model initiates its retrieval procedure. Long-term memory strength for each element in LS was computed by summing up its self-connection and the strength of its connections to other elements. If the learning mechanism of LS is correct, then it should be able to build an appropriate offline episodic memory representation from online activation levels. Consequently, this memory representation of LS should be related to the likelihood of human recall for corresponding elements.

The first measure, recall proportion, is necessary because it is evaluating the model's quantitative performance, i.e., how much does the model predict human recall. This measure, however, does not take into account response variability, namely it does not distinguish a concept recalled by only one reader from another concept recalled by all readers. The reason is that LS's recall is deterministic, therefore concepts are classified in a dichotomous manner. Unlike the first measure, the second measure takes into consideration variability of readers' responses. It mimics the quality of the model's performance because a high correlation of memory strength with human recall is an indicator that the model produces a similar memory pattern as human readers. This measure, however, does not tell the quantitative aspect of the model's performance because a high correlation in memory strength could be based on only a small recall proportion. Therefore, these two measures are complementary and are both necessary for evaluating the model's performance.

Results

The results of this study are described in two sections. The first section describes the effects of various cohort sizes and learning rates on LS's recall proportion. The second section describes the effects of cohort sizes and learning rates on LS's memory strengths.

The effects of parameter values in determining LS's prediction of recall proportion. For the Epaminondas (Epmin) story, Table 2 shows that the proportion of recall predicted by LS depends on the values of the cohort parameters, $F(5, 25) = 30.62$, $p < .0001$. The average proportion of recall predicted by LS jumps from .49 to .94 and .97 as cohort value increases from .00 to .03, and .06; it reaches the peak value 1.0 when the cohort value is equal or larger than .09. It appears that LS can predict a very high proportion of recall under certain cohort parameter values.

Overall, LS's ability to predict recall proportion does not vary with competitive learning rate, $F(5, 25) = 2.04$, $p > .10$. This pattern does not necessarily discount the existence of an effect of learning rate. Since there is no within cell variability in Table 2, a statistical analysis of the interaction of the effects of these two parameters can not be conducted. Nonetheless, the individual cells of Table 2 indicate that there is an effect of learning rate when cohort sizes are small. When the cohort parameter is zero, recall proportion predicted by LS drops from .8 to .36 when learning rate increases from .0 to .6 and remains at the same low proportion through the learning rate value of 1.5. The predicted recall proportion decreases from 1.0 to .8 as learning rate increases from .00 to 1.5 when cohort was .03. A roughly similar pattern is discovered for cohort value .06. There are no signs of the effect of learning rate for cohort values larger than .09. These suggested that learning rate played a role in determining LS's prediction of recall proportion under lower cohort values and played little or no role under higher cohort values.

For the Tiger story, Table 2 indicates that LS's ability to predict recall proportion does not vary with cohort values, $F(5, 25) = 2.15$, $p > .05$. The average recall proportion accounted for by LS for this story consistently is high, ranging from .85 to .96. However, it never reaches 1.0 due to over-recall of the model resulting in some penalty.

As for the Epamin story, LS's ability to predict recall proportion is not affected by learning rate for the Tiger story, $F(5, 25) < 1$. On average, the effect of learning rate in determining recall proportion for this story shows little variation (between .91 and .96). No statistical analysis on the interaction of these two parameters can be conducted because there is no within cell variation. Individual cells in Table 2 suggest that learning rate has differential effects for some cohort values. LS predicts .63 recall proportion when learning rate is .00 and .30 under the zero cohort condition; its prediction then soars to .96 for learning rates larger than .60. An opposite pattern is observed for cohort value at .03. LS predicts .96 recall proportion when learning rate is under .60 and drops to .83 when learning rate is higher than .60. There is no impact of learning rate when cohort is higher than .06. These results are similar to those for the Epamin story in that learning rate does not affect LS's prediction of recall proportion when cohort values are high but affects the models' performance when cohort values are low.

The effects of parameter values in determining LS's prediction of memory strength. For the Epamin story, the results in Table 3 shows that cohort values affect LS's correlation of long-term memory strength with human recall, $F(5, 25) = 75.68$, $p < .001$. On average, LS correlates with human LTM significantly only when cohort equals .00 and .03 (average correlation coefficients are .52 and .54 respectively, ps < .05). LS does not have a significant correlation with human recall if cohort values are larger than .06. It appears that LS with low cohort values can account human memory strength better than LS with high cohort.

The performance of LS is not affected by the value of the learning rate parameter, $F(5, 25) = 1.16$, $p > .05$. On average, learning rate appears to cause some fluctuations of correlation, but the effects are too small to have a significant contribution. Note that the correlation of the model's long-term memory strength and human recall is essentially zero under high cohort conditions for all values of learning rate. It is impossible to conduct a statistical analysis on the interaction of cohort and learning rate parameters because there is no within cell variation in Table 3. A careful inspection of individual cells in Table 3 for Epamin reveals very homogeneous correlation sizes within each cohort value across different learning rates. This suggests that the cohort parameter dominates the correlation of long-term memory strength between LS and human recall whereas learning rate plays a non-significant role.

For the Tiger story, a similar but not identical pattern is observed as in the Epamin story. Different cohort values have strong effects on the performance of LS, $F(5, 25) = 23.4$, $p < .001$. LS long-term memory strengths are significantly correlated with human recall when cohort values are .00, .03, and .06 (average correlation coefficients were .57, .59, and, .45 respectively, ps < .05). LS does not appear to have any similarity to human recall when cohort is larger than .09 (all ps > .05). It appears that LS with low cohort values predicts human memory strength whereas LS with high cohort values could not.

The size of the learning rate parameter does not have a significant effect on LS's long-term memory strength for the Tiger story, $F(5, 25) < 1$. This lack of effect of learning rate indicates that LS's correlation of long-term memory strength with human recall is entirely determined by cohort values and this is consistent with the pattern for the Epamin story. Individual cells do not show any variations neither suggesting no interaction effect.

Discussion

The results of this study suggest that cohort and learning rate parameter play different roles in the LS model. This is supported by the fact that LS's performance is affected both by cohort and learning rate parameters indicating the model has correctly included these two different parameters as two separate constructs. Additionally, LS can predict a high proportion of human recall and its long-term memory strength correlation with human recall is significant under some parameter values. This will provide a proper starting point for further investigating different features of this model.

Although LS's performance is affected by the two parameters, its dependency on them is not the same. The performance of LS on recall proportion appears to be dependent on cohort values but only marginally on the learning rate at some conditions. The dependence on cohort is readily interpretable. As cohort size increases, reading elements are more likely to connect with each other and the memory representation is more richly interconnected. Consequently, it is less likely to reach a dead end during retrieval and the amount of recall is usually large. The model therefore predicts a rather high proportion of human recall. If cohort size is small, there are smaller amount of spread of activation in the model resulting fewer connections among nodes. Consequently, the model is more likely to encounter a dead end during retrieval. Therefore it tends to predict a small amount of human recall.

The model's performance on memory strength also is dependent more on cohort size. The larger the cohort sizes the lower the model's predictive power of human memory strength. This is the opposite pattern as the model's performance on recall proportion. When the cohort is large, reading elements that appear early in the text have a large advantage because once they are read and become part of a cohort, they are activated repeatedly due to large effect of cohort activation. As a result, the model's memory strength for reading elements does not reflect their relative importance in the text, instead it is more determined by order of input (van den Broek, et al., 1999). Thus, it is no surprise to observe a low correlation between LS's memory strength and that of the participants under high cohort conditions. Under conditions of small cohort size, the effect of cohort activation does not obscure the relative importance of elements in the text. Therefore LS's memory strength is more closely related to human memory strength.

The model's performance, both on recall proportion and memory strength, appears relatively unaffected by the competitive learning rate. Although the learning rate affects LS's performance under certain conditions, its overall impact is quite minor. The reason for this low effect of learning rate is probably due to the small amount of learning effect that can occur in the model's architecture. More will be discussed in general discussion.

With these results, one can select the best-fit parameter values to maximize the model's performance. There are two criteria for selecting the parameter values. First, the model needs to predict at least 80% of the human recall proportion. Second, the model needs to show high correlation with human recall in its long-term memory strength. The first criterion is necessary because a good model should be able to predict a certain amount of <u>quantity</u> of human recall. An 80% threshold is reasonable especially in the light of the fact that the number of reading elements for each story used in this study was usually no more than thirty. If a selected version of LS predicts 80% of human recall for a certain story with thirty reading elements, it suggests that the model recalls twenty-four elements, which would allow minimally sufficient degrees of freedom for computing the correlation of memory strength between LS and human recall. If the recall proportion threshold is lower than 80%, then the selected LSs would recall few reading elements. As a result, the model may not have enough degrees of freedom to reliably compute long-term memory strength correlations with human recall. The second criterion is to ensure the model predicts the quality of human recall. If selected parameter values were correct, the model should be able to mimic human memory strength. Therefore they should have significant correlations with human data. The selection of appropriate parameter values thus was to balance between the two measures of model performance. Cohort and learning rate values that enable LS meet these two criteria simultaneously from study 1-a were selected. The same parameter values then would be applied to a new set of stories.

By applying this two-step procedure, zero cohort conditions were excluded because they predict very low recall proportion for most conditions with an average of .49 for the Epamin story (except for the cell of zero cohort zero learning rate, see Table 2 & 3). Even though its predicted recall proportion was sufficiently high for the Tiger story (except for .00 and .30 learning rate under zero cohort), the performance on these two stories counteract with each other. By applying the second step of the selection procedure, cohort values equal .06 and above were excluded because their long-term memory strengths did not significantly correlate with human recall for both stories. Although at cohort equals .06 LS correlated significantly with human data for the Tiger story, it failed to do so for the Epamin story. Therefore it did not meet the criterion. Consequently, conditions with cohort value .03 met all the requirements with equal plausibility for all six learning rates for both stories. A conservative middle value of learning rate .9 was chosen to maximize the long-term stability of the model.

Study 1-b

The purpose of this study was to cross-validate the parameter values selected from study 1-a and to test how robust these parameter values are for another set of stories. With the "best" parameter values from study 1-a, one could ask how general these values are and how robust of the model's performance would be. Did these values only work well for these two stories or would they work for other stories? It is entirely possible that there are some unique features in these two stories made LS worked well. The aim of this study was to further examine the performance of LS on two additional stories using precisely the same parameter values derived for the two stories in study 1-a.

Method

Materials. Two new stories, Judy's Birthday (Judy) and Fox and Bear (Fox), from Trabasso et al. (1984) were adopted. Their parsing and causal analysis were conducted using the same principle as in study 1-a.

Participants. The participants are the same as in study 1-a.

Procedure. The procedures for collecting data and scoring were the same as in study 1-a. Cohort value of .03 and competitive learning rate .9 were adopted based on the earlier analysis and were applied here to model the two new stories. These two runs were testing the stability of the parameter values and were evaluating the global performance of LS. Two additional sets of simulations were carried out to test more specific assumptions of LS. The first additional set was to test the necessity of causal and referential inferences in the model. This was achieved by assigning zero activation values for concepts activated through one or both processes. The second additional set was achieved by equating all activation values to three or by reversing activation values for the different sources. These analyses were conducted only on the two new stories to avoid possible logical circularity. The same two dependent measures in study 1-a were adopted to evaluate the success of each version of the model, namely recall proportion and the correlation of memory strength between LS and human recall.

Results

Performance on recall proportion. As can be seen in Figure 1, LS predicts 89% of human recall for the Judy story. For the Fox story, LS predicts 85% of human recall. These two proportions are within the range of 83% (Tiger story) and 96% (Epamin story) observed in the first study. This shows that LS can account for a very high recall proportion and that this capability is quite stable across different stories.

Performance on correlation with human memory strength. As can be seen in Figure 2, the correlation of the memory strengths for LS and human recall was .50 and .47 (*p*s < .05), for the Judy story and the Fox story, respectively. These two values were very close to the correlation values for the other two stories (.58 and .53 respectively). Again this indicates that LS produces a stable memory representation after reading. As a result, the same parameter values generate strikingly similar patterns of results for different stories.

Discussion

The results of this study confirmed that LS is a valid model of comprehension because it could capture substantial amount of variance in human comprehension performance. Furthermore, the fact that its performance for new stories was very similar to that for old stories indicated that its performance was reliable and not simply due to the particular stories. This reliability could be attributed to the overall structure of the model, which could tolerate unique text structures of different stories. The results also show that the cohort and learning rate parameters are not redundant. However, the cohort parameter has played a much more dominant role in determining the model's performance. These findings are very useful in at least two ways: first, LS is not only valid but also reliable; second, potential users of LS know reasonable starting parameter values.

Study 2

In addition to the two parameters tested in Study 1-a and 1-b, there are other crucial parameters in LS that are worthy of exploration. One likely candidate is the current scheme of assigning activation values for different sources of activation in the model. Although the relative activation for direct text input, causal and anaphoric inferences are informed by empirical data, it is only a best guess. One great advantage of a computational model is that it allows one to test different components of the model by disabling them. For example, setting zero activation values for referential and causal inferences would test the assumptions that they were central online processes during reading. In this study, two alternative activation schemes were implemented in LS to serve as control groups and compared to model's performance at its default activation values. Therefore this would constitute a test of the plausibility of the original weightings for different sources of activation.

Method

Materials. The same Judy and Fox stories in Study 1 were used.

Participants. Recall data from Study 1 for Judy and Fox stories were used.

Procedure. To set up simulations for different activation schemes, two new schemes were created by changing relative activation values for each source. The first new scheme has three variations: (1) assigning zero activation values for all causal inferences; (2) assigning zero activation values for all anaphoric inferences; (3) assigning zero activation values for all causal and anaphoric inferences simultaneously. The second new scheme also has two variations: (1) assigning all activation sources with equally low activation value of 3; (2) assigning reverse relative activation values for four major sources of activation: five for enabling, four for referential and causal connections, and three for mention (instead of five for mention, four for referential and causal connections, and three for enabling in the model's default activation scheme). The first variation provides a baseline to examine whether extra activation values in addition to the value of 3 for mention, referential and causal connections are necessary in the model. The second variation serves the same purpose as the first one except taking a more extreme form. The memory strengths of these simulations were then correlated with recall proportions of human data. All these reduced or distorted LS models were contrasted with original full model performance. The procedures for running these two simulations were the same as in Study 1.

Results

Figure 3 shows the proportion of recall that LS predicts when causality and/or referential coherence are removed as sources of activation. LS's predictive power was reduced substantially and fell below 80% for all the modified versions. As a result, none of the modified LSs were eligible for computing LTM correlation with human recall. The results of this set of analysis demonstrate that causal and referential connections are necessary for LS and that removal of either one reduces the model's performance substantially.

The patterns in Figure 3 also indicate that LS versions with new schemes of assigning activation values to different sources did not account for substantial proportions of recall. Consequently, both modified versions of LS did not meet the criteria for viable implementation. The results suggest that the current rule of assigning activation within LS is reasonable. Changing the rule of allocating activation damages the model's performance to a very substantial degree.

Discussion

The simulations in this study clearly showed that both causal and referential connections are essential for the model's performance. Once either of them was taken away, LS simply could not predict human performance. This is a direct support for the necessity of these two online comprehension mechanisms. The modeling results that causal and referential connections are fundamental to comprehension are totally in line with behavior studies.

The modeling data on how LS assigns activation values also indicate that it is reasonable to have the current scheme of assigning activation values. More specifically, it appears reasonable to assign extra activation for reading elements from the explicit text input and somewhat less for causal and referential inferences, and even less for enabling connections which is considered a weaker form of causality. Modification of this scheme reduces LS's performance substantially. Of course there are numerous ways of assigning activation values to sources of activation and we have only tested a limited set of them. Simulations with other activation schemes along with possible empirical research would further test the validity of the current assignment procedure.

General Discussion

Any scientific model strives for explanatory adequacy. Models in psychological science should be able to account for human performance and they should have psychological plausibility. In a domain as complex as comprehension, this research has shown that a computational model such as the Landscape model is fruitful for tracking sub-processes of comprehension and components of the model.

One major goal of this research is to validate a computational model of comprehension, the Landscape model (LS). The model is based on the notion that readers routinely generate referential and causal inferences during reading to attain coherence from the text. The end result of comprehension is a memory representation of the situations that the text refers to. LS treats the transformation of online activation to offline representation as a learning phenomenon and uses a formal learning rule to capture the relations.

The purpose of study 1 was to test whether two important parameters, the cohort parameter and the competitive learning rate parameter, affect LS's performance. If LS has construct validity, different values of the cohort parameter and the competitive learning rate will influence its performance. Additionally and very importantly, the model should be able to predict a substantial amount of human performance for some range of its parameter values.

The results of study 1-a support the construct validity of LS. Several patterns are consistent with this conclusion. First, the model's prediction of recall proportion, the ratio of model recall over human recall, was very high. Interestingly, the model's prediction was affected largely by the cohort parameter and to a less extent by the competitive learning rate. Second, there were considerable similarities between the model and human memory strengths; These again were mostly determined by the cohort parameter and less affected by the learning rate. Third, the model accounted for a very high recall proportion when the cohort parameter was medium or high and it significantly predicted human memory strength when the cohort parameter was medium or low. In other words, LS's performance is affected by both parameters despite the fact that one of them is more dominant than the other. Additionally, LS predicts human performance very well with medium cohort parameter values. Therefore, LS's validity is supported. Moreover, both the cohort and learning rate parameters are psychologically valid in the context of reading comprehension, adding another piece of evidence to support LS.

Study 1-b further tests the validity of LS by examining its performance for other stories. By applying the best parameter values selected from two stories used in study 1-a, study 1-b is also testing the reliability of the model's parameters. The results showed that LS's performance on new stories was similar to that for previous two stories, both for recall proportion and memory strength correlation with human memory. Therefore, study 1-b supports the validity of LS. The model's successful performance appears due to its structure, not merely a lucky selection of stories per se.

Two additional results from study 1-b justify two crucial assumptions of LS. The first additional set of simulations tested LS's assumption that referential and causal inferences are essential for comprehension. The results indicate that removal of inferential and causal connections from the model substantially decreased its performance. Thus LS predicts that readers reliably make these two types of inferences, consistent with human comprehension data. The second additional set of simulations examined the model's assignment of activation values for different sources. LS currently assigns a high activation value for explicit text input, a medium value for referential and causal connections, and a low value for enabling connections. Two alternative methods of assigning activation were created by (1) assigning equal low values for all sources and (2) reversing the high and low activation values for their respective sources. The model's performance for these alternative methods was substantially damaged, indicating that the model's current method of assigning activation values is reasonable.

In summary, these studies have demonstrated the validity of LS. Together they imply that as readers go through a text, they do rely on referential and causal inferences to build a coherent representation of what they read from the text. Later on, they use this representation for recall. This research shows that LS is effective in capturing how these online activations are transformed into this stable offline representation.

The success of LS has demonstrated that a computational approach is fruitful for capturing the processes and results of comprehension. Meanwhile, it raises several important issues for further consideration. The results in study 1 indicate that the cohort parameter influences the model's fit more than does the learning rate parameter. Specifically, the effects of the learning rate only occurred when the cohort parameter was very low for recall proportion and when the cohort was high for memory strength patterns. This pattern supports the model's postulations of the distinction of the cohort and the learning rate parameters. At the same time, it also raises an interesting question of why the learning rate is less important than the cohort parameter.

There are two possible reasons for the relative importance of these two parameters. One possibility is that it is due to the current particular implementation of LS, which used a modified delta rule to reduce the amount of expectancy for each element. Reducing the amount of expectancy lessened the asymmetry in episodic connections in the model and greatly eliminated negative connections, which are less clear for interpretations. The procedure of reducing the amount of expectancy might have changed the validity of the learning rate. Another possibility is that the cohort parameter is a more important determinant of human behavior than the competitive learning rate. The cohort parameter controls the number of possible active elements within one cycle and the possible frequency of being activated for particular reading elements. A large cohort value will make the model hold many elements for many cycles. Human cognition may be sensitive to this property of high frequency of co-occurrence among reading elements. In contrast, a high learning rate will make the model form a large expectancy and update the connections cautiously. This property may not be important for forming an episodic memory through narrative comprehension. Future research specifically focused on properties of these two parameters will help to clarify this issue.

A very important factor that determines the fitting of parameter values is the choice of dependent measures. The choices of distinct dependent measures may alter the models' performance, thereby changing the parameter fitting values. The fact that high cohort values increase the model's prediction of recall proportion and low cohort values benefit its prediction of human memory strength suggests that these two measures may not be assessing the same aspect of the model's performance. Upon close examination these two measures (recall proportions and memory strength correlation) indeed appear to differ in nature. For instance, the correlation of the model's memory strength with human data takes into account readers' response variability because human memory strength is determined by the number of readers who recall a certain element; the more people recall it the stronger the memory strength. It is a <u>qualitative</u> indicator of the model, i.e., how similar is the model's memory strength to human data. In contrast, recall proportion does not take into account readers' response variability. It dichotomizes the reading elements: non-recalled elements or recalled elements regardless the number of people has recalled them. It is a <u>quantitative</u> indicator of the model, i.e., how much proportion of human recall does the model predict. These two measures seem to be complementary.

An issue here is how the use of multiple measures influences the choice of parameter values. If the selection of proper parameter values is solely based on the model's performance on recall proportion, most cohort values will fit except for zero cohorts because zero cohort conditions do not predict high enough recall proportions (see Table 2). In contrast, if the selection process is based on the model's memory strength, only low to zero cohorts will fit because their correlations with human memory strengths are significant (see Table 3). However, this research considered the model's performance on these two measures simultaneously and a low cohort (but not zero) medium learning rate condition was selected. This analysis suggests that the use of multiple dependent measures influence the selection of parameter values and it seems to provide more constraints on the selection of parameters, especially if these measures are complementary.

Another issue is what we gain by using multiple dependent measures to select parameter values. The results from study 1 suggest that we could have chosen medium or high cohort values if we used recall proportion as a dependent measure and zero cohort for the memory strength measure. In either case, we could have reached the false conclusion that the model is valid if cohort is medium to high (for the recall proportion measure), or that the model is valid if cohort is zero (for the memory strength measure). By applying multiple measures, the selection of parameter values is more specific and the evaluation more appropriate. Therefore, we can only conclude that the model is valid with low cohort values (for both measures). With proper complementary measures, the evaluation of LS or any computational models can be enhanced. This research has carefully and successfully used two complementary measures to evaluate LS. Additional dependent measures could be used to assess other aspects of the model and provides an even more thorough evaluation of LS.

Conclusion

This research has made several contributions to the study of language comprehension and human memory. The research has provided more evidence for the LS model. The evidence is partially from the close correspondences between human recall and model memory representation. A major contribution of this research is that we were able to cross validate the performance of the model beyond specific parameter values. Furthermore, to our knowledge, this is the first research to systematically investigate the inner workings of comprehension by simulations. Together, these studies have provided insights and generated research ideas for language comprehension.

One major line of future research should be to test more aspects of LS. For example, one can evaluate the model's performance using more constrained indicators, such as recall order to provide more stringent tests of the model. Another possible line of future research is to test online aspects of the model. Since LS produces asymmetric connections among reading elements, it would be informative to test this specific aspect of the model. This will allow us to compare the model with other models such as the influential Construction-Integration model (Kintsch, 1988). The Construction-Integration model, in contrast to LS, has assumed that memory representation of comprehension is symmetrical. A third line of future research is to add more "features" to LS. For example, working memory has been shown to influence language comprehension (Just & Carpenter, 1992). Therefore, adding working memory capacity constraint on LS is desirable and it has been implemented in the Windows version of the LS program (Tzeng, et al., 2005). Whether working memory plays a similar role in this model remains to be tested. Adding a semantic component to the model will be a challenging but very important task. It is important because semantic memory plays a central in comprehension. It is challenging because it is difficult to capture reader's background knowledge in a computational manner. The advent of Latent Semantic Analysis (Landauer & Dumais, 1997) marks a new possibility to approximate readers' background knowledge. A precaution here is that adding more features generally increases the complexity of a model which damages parsimony. That is one reason for the LS model to implement a retrieval module without adding explicit theories such SAM (Gillund & Shiffrin, 1984). The current retrieval mechanism actually mimics the retrieval processes of SAM without adding much complexity to the model. It is possible LS has lost a great deal of predictive power by not incorporating models such as SAM. This remains an open question.

It is important to point out that there are several limitations of this research. There were only a small number of participants and an even smaller number of stories included in this research. This would greatly limit the generality of our conclusions. Certainly, there are always limitations like these in most research. However, it is always desirable to have enough sampling of items and subjects. The choice of dependent measures is another issue for this research. The recall proportion and recall ratio per story derived from human recall appears to be very sample dependent, both depending on sizes and characteristics of stories and participants. This echoes the limitations that were raised previously. If different sets of stories or participants had been chosen, these two measures might have been very different. As a result, the performance of the model might have been changed. It is crucial to devise some dependent measures that can cope with this issue. Finally, the lack of a standardized procedure of model testing is another concern not only for comprehension modeling but also in other areas as well. In some studies, we were able to use correlation coefficients as indicators, which is the most formal method of this research so far. Other times, we even resort to subjective judgments to decide whether .80 is high enough, for instance. This is certainly unsatisfactory from a methodological point of view. Modeling research needs to be more informed in this aspect.

This research has shown that a computational approach is feasible not only for investigating low level of cognition, it is also useful even for a domain as complex as comprehension. By keeping track of sub-processes of a model, one is not only able to scrutinize each component individually but also oversees the dynamics of the whole. By comparing the predictions of a cognitive model with performance of human participants, we are on a right track of unfolding human cognition.

References

Anderson, J. R. (1995). *Learning and memory: An integrated approach*. NY, John Wiley & Sons, Inc.

Barto, A. G. (1990). Connectionist learning for control: An overview. In W. T. Miller, III, R. S. Sutton, & P. J. Werbos (Eds.), *Neural network for control*. Cambridge, MA: MIT Press.

Bloom, C. P., Fletcher, C. R. van den Broek, P., Reitz, L., & Shapiro, B. P. (1990). An online assessment of causal reasoning during comprehension. *Memory & Cognition*, *18*, 65-71.

Dell, G. S., McKoon, G., & Ratcliff, R. (1983). The activation of antecedent information during the processing of anaphoric reference in reading. *Journal of Verbal Learning and Verbal Behavior*, *22*, 121-132.

Dopkins, S., Klin, C., & Myers, J. L. (1993). Accessibility of information about goals during the processing of narrative texts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 70-80.

Duffy, S. A., & Rayner, K. (1990). Eye movement and anaphor resolution: Effects of antecedent typicality and distance. *Language and Speech*, *33*, 103-119.

Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movement and immediacy of processing. *Journal of Verbal Learning and Verbal Behavior*, *22*, 75-87.

Fletcher, C. R. (1994). Levels of representation in memory for discourse. Comprehension and memory of narratives texts: Inferences and coherence. In M. A. Gernbacher (Ed.), *Handbook of Psycholinguistics* (pp. 589-607). New York: Academic Press.

Garrod, S., & Sanford, A. (1990). Referential processes in reading: Focusing on roles and individuals. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 465-484). New York, N.Y.: Academic Press.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, *32*, 99-156.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.

Gernsbacher, M. A. & Faust, M. E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 245-262.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1-67.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, *117* (3), 227-247.

Goldman, S. R., Golden, R., & van den Broek, P. (in press). Why are computational models of text comprehension useful? In F. Schmalhofer & C. Perfetti (Eds), *Higher level language processes in the brain*. Mahwah, NJ: Erlbaum.

Graesser, A. C., Olde, B., & Klettke, B. (2005). How does the mind construct and represent stories? In M. Greene, J. Strange, & Brock, T. (Eds.), *Narrative Impact* (pp.1-51). Mahwah, NJ: LEA.

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, *13*, 512-521.

Just, A. M., & Carpenter P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122-149.

Kendeou, P., & van den Broek, P. (2005). The effects of readers' misconceptions on comprehension of scientific text. *Journal of Educational Psychology*, *97*, 235-245.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: Construction-integration model. *Psychological Review*, *95*, 163-182.

Klin, C. M., & Myers, J. L. (1993). Reinstatement of causal information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. *19*, 554-560.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211-240.

Linderholm, T., Virtue, S., Tzeng, Y., van den Broek, P. (2004). Fluctuations in the availability of information during reading: Capturing cognitive processes using the landscape model. *Discourse Processes*. *37*(2), 165-186.

McClelland, J. L., & Rumelhart, D. L. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*, 159-188.

McKoon, R. & Ratcliff, G. (1992). Inferences during reading. *Psychological Review*, *99*, 440-466.

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, *117*(3), 363-386.

Singer, M. (1994). Inference generation during reading. In M. A. Gernbacher (Ed.) *Handbook of Psycholinguistics* (pp. 479-515). New York: Academic Press.

Singer, M., Halldorson, M., Lear, J. C., & Andrusiak, P. (1992). Validation of causal bridging inferences in discourse understanding. *Journal of Memory and Language*, *31*, 507-524.

Thurlow, R E. (1991). *The inference of causal antecedents during the reading of narratives*. Unpublished doctoral dissertation, University of Minnesota.

Trabasso, T., Secco, T., & van den Broek, P. W. (1984). Causal cohesion and story coherence. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.). *Learning and comprehension of text* (pp. 83-111), Hillsdale, N.J.: Lawrence Erlbaum Asociates.

Trabasso, T., van den Broek, P. W., & Suh, S. Y. (1989). Logical necessity and transivity of causal relations in stories. *Discourse Processes*, *12*, 1-25.

Tzeng, Y., van den Broek, P., Kendeou, P., & Lee, C. (2005). The computational implementation of the Landscape model: Modeling inferential processes and memory representations of text comprehension. *Behavior Research Methods*, *37*(2), 277-286.

van den Broek, P. W. (1990). The causal inference maker: Towards a process model of inference generation in text comprehension. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 423-445). New York, N.Y.: Academic Press.

van den Broek, P. W. (1994). Comprehension and memory of narratives texts: Inferences and coherence. In M. A. Gernbacher (Ed.) *Handbook of Psycholinguistics* (pp. 539-588). New York: Academic Press.

van den Broek, P. W., & Lorch, R. F., Jr. (1993). Network representations of causal relations in memory for narrative texts: Evidence from primed recognition. *Discourse Processes*, *16*, 75-98.

van den Broek, P. W., Risden, K., Fletcher, C. R., & Thurlow, R. (1996). A ''landscape'' view of reading: Fluctuating patterns of activation and the construction of a stable memory representation. In B. K. Britton. & A. C. Graesser (Eds.), *Models of understanding text* (pp. 165-187). Hillsdale, N.J.: Erlbaum.

van den Broek, P. W., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading: Inferences and the on-line construction of a memory representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading*, Mahwah, NJ:Erlbaum.

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

Wang, H., Johnson, R. J., & Zhang, J. (1998) UEcho: A model of uncertainty management in human abductive reasoning. *In Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 1113-1118). Hillsdale, NJ: Lawrence Erlbaum Associates.

Zhang, J., Johnson, T. R., & Wang, H. (1996). Order effects and frequency learning in belief updating. *In Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Appendix 1 Stories materials

Title: Tiger

Once there was a poor, old woman.
Together with her husband she lived in the forest.
One day she was walking in the hills.
She came upon the entrance to a tiger's cave.
She wanted one of the tiger's whiskers.
She wanted to make a medicine of the whisker for her husband.
She put a bowl of food in front of the entrance to the cave.
She softly sang a song which put her the tiger to sleep.
The old Woman quickly pulled out one of the whiskers.
Very quickly, she ran down the hill.
Panting, she came home.
She was very pleased.

Title: Epaminondas

Once there was a little boy.
The little boy lived in a hot country.
The little boy's mother told him to take some cake to his grandmother.
The little boy's mother warned him to hold the cake carefully.
The cake shouldn't break into crumbs.
The little boy put the cake in a leaf under his arm.
The little boy carried the cake to his grandmother's house.
The little boy got to his grandmother's house.
The cake has crumbled into tiny pieces.
The little boy's grandmother told him to that he was a silly boy.
He should have carried the cake on top of his head.
The cake would not have broken into pieces.
The little boy's grandmother gave him some butter to take back to his mother.
The little boy wanted to be very careful with the butter.
The little boy put the butter on top of his head.
The little boy carried the butter home.
The sun was shining hard.
The little boy got home.
The butter had all melted.
The little boy's mother told him that he was a silly boy.
The little boy should have put the butter in leaf.
The butter would have gotten home safe and sound.

Title: Fox and Bear

There was a fox and a bear who were friends.
One day they decided to catch a chicken for supper.
They decided to go together because neither one wanted to be left alone and they both liked fried chicken.
They waited until night time.

Then they ran very quickly to a nearby farm where they knew chickens lived.
The bear, who felt very lazy climbed up on the roof to watch.
The fox then opened the door of the henhouse very carefully.
He grabbed a chicken and killed it.
As he was carrying it out the henhouse, the weight of the bear on the roof caused the roof to crack.
The fox heard the noise and was frightened but it was too late to run out.
The roof and the bear fell in killing five of the chicken.
The fox and the bear were trapped in the broken henhouse.
Soon the farmer came out to see what was the matter.

Title: Judy's Birthday
Judy is going to have a birthday party.
She is ten years old.
She wants a hammer and a saw for presents.
Then she could make a coat rack and fix her dollhouse.
She asked her father to get them for her.
Her father did not want to get them for her.
He did not think that girls should play with a hammer and a saw
but he wanted to get her something.
So he bought her a beautiful new dress.
Judy liked the dress
but she still wanted the hammer and the saw.
Later she told her grandmother about her wish.
Her grandmother knew that Judy really wanted a hammer and a saw.
She decided to get them for her
because when Judy grows up and becomes a woman
she will have to fix things when they break.
Then her grandmother went out that very day and bought to tools for Judy.
She gave them to Judy that night.
Judy was very happy.
Now she could build things with her hammer and saw.

Appendix 2 Input matrix for Tiger story

	Reading Cycle											
	1	2	3	4	5	6	7	8	9	10	11	12
poor	5	3	0	0	0	0	0	0	0	0	0	0
old	5	3	0	0	0	0	0	0	5	0	0	0
woman	5	4	4	4	4	4	4	4	5	4	4	4
live	0	5	3	0	0	0	0	0	0	0	0	0
forest	0	5	3	0	0	0	0	0	0	0	0	0
husband	0	5	0	0	0	5	0	0	0	3	0	0
walk	0	0	5	3	0	0	0	0	0	3	0	0
hill	0	0	5	3	0	0	0	0	0	5	0	0
come	0	0	0	5	4	3	0	0	0	0	0	0
entrance	0	0	0	5	4	3	5	0	0	0	0	0
tiger	0	0	0	0	5	3	0	5	0	0	0	0
cave	0	0	0	0	4	3	5	0	0	0	0	0
wantmake	0	0	0	0	5	5	4	0	4	0	0	0
whiskers	0	0	0	0	5	5	4	0	5	4	0	4
medicine	0	0	0	0	0	5	4	4	4	0	0	0
put	0	0	0	0	0	0	5	3	0	0	0	0
bowlfood	0	0	0	0	0	0	5	3	0	0	0	0
front	0	0	0	0	0	0	5	3	0	0	0	0
sing	0	0	0	0	0	0	0	5	3	0	0	0
sleep	0	0	0	0	0	0	0	5	3	0	0	0
pull	0	0	0	0	0	0	0	0	5	4	0	4
run	0	0	0	0	0	0	0	0	0	5	4	0
home	0	0	0	0	0	0	0	0	0	0	5	4
pant	0	0	0	0	0	0	0	0	0	0	5	4
pleased	0	0	0	0	0	0	0	0	0	0	0	5

Table 1
Number of word, sentence, mean and standard deviation recall proportion, and recall ratio for each story

	TIGER	EPMIN	JUDY	FOX
number of sentences	12	22	20	13
number of words	100	202	174	160
mean recall proportion	0.24	0.53	0.48	0.34
standard deviation of mean recall proportion	0.16	0.22	0.24	0.25
recall ratio per story	0.92	1	0.96	0.93

Table 2
Recall Proportion as a Function of Values of Cohort and Competitive Learning Rate for Epamin and Tiger Stories

	Cohort													
	Epamin							Tiger						
Learning Rate	0	0.03	0.06	0.09	0.12	0.3	Mean	0	0.03	0.06	0.09	0.12	0.3	Mean
0	0.8	1	1	1	1	1	0.97	0.63	0.96	0.96	0.96	0.96	0.96	0.91
0.3	0.72	1	1	1	1	1	0.95	0.63	0.96	0.96	0.96	0.96	0.96	0.91
0.6	0.36	0.92	1	1	1	1	0.88	0.96	0.96	0.96	0.96	0.96	0.96	0.96
0.9	0.36	0.96	1	1	1	1	0.89	0.96	0.83	0.96	0.96	0.96	0.96	0.94
1.2	0.36	0.96	1	1	1	1	0.89	0.96	0.83	0.96	0.96	0.96	0.96	0.94
1.5	0.36	0.8	0.8	1	1	1	0.83	0.96	0.83	0.96	0.96	0.96	0.96	0.94
Mean	0.49	0.94	0.97	1	1	1	0.9	0.85	0.9	0.96	0.96	0.96	0.96	0.94

Table 3
Correlation of Memory Strength between LS and Human as a Function of Values of Cohort and Competitive Learning Rate for Epamin and Tiger Stories

	Cohort													
	Epamin							Tiger						
Learning Rate	0	0.03	0.06	0.09	0.12	0.3	Mean	0	0.03	0.06	0.09	0.12	0.3	Mean
0	0.52	0.53	0.27	0.18	0.15	0.13	0.3	0.58	0.59	0.42	0.19	0.11	0.08	0.33
0.3	0.52	0.57	0.3	0.17	0.11	0.08	0.3	0.57	0.59	0.43	0.21	0.09	0.05	0.32
0.6	0.52	0.53	0.31	0.17	0.09	0.05	0.28	0.57	0.59	0.44	0.22	0.08	0.03	0.32
0.9	0.52	0.53	0.32	0.18	0.09	0.31	0.33	0.57	0.58	0.44	0.25	0.07	0.01	0.32
1.2	0.51	0.53	0.33	0.19	0.09	0.31	0.33	0.57	0.58	0.47	0.26	0.06	0	0.32
1.5	0.51	0.53	0.35	0.2	0.34	0.3	0.37	0.57	0.58	0.49	0.27	0.05	0	0.33
Mean	0.52	0.54	0.31	0.18	0.15	0.2	0.32	0.57	0.59	0.45	0.23	0.08	0.03	0.32

49-3-A-061106英245-269 9/10/2007 1:06 PM 頁面 267

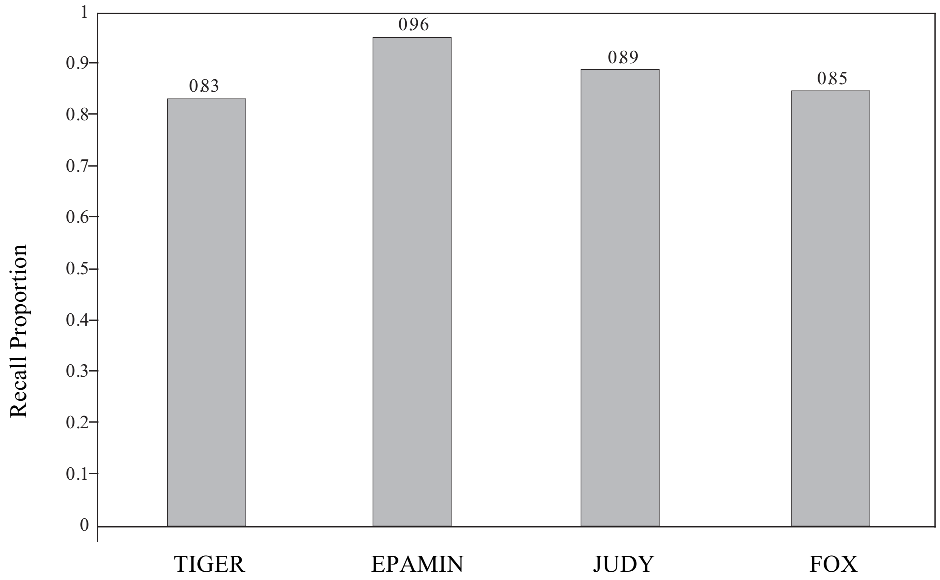

Figure 1. Proportion of Recall Predicted by LS as a Function of Story

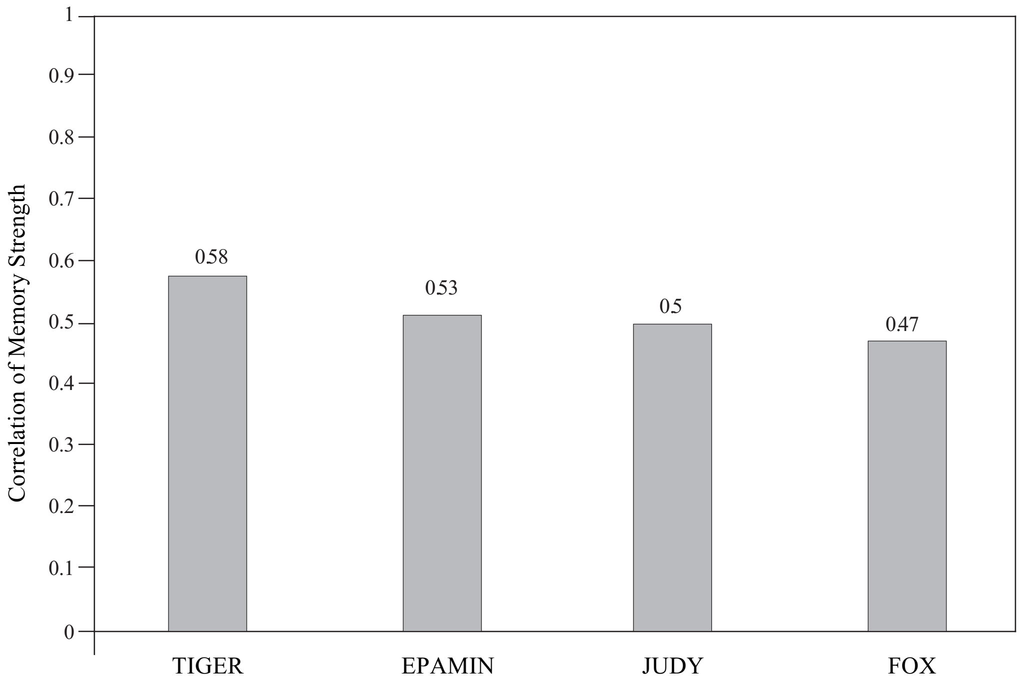

Figure 2. Correlation of Memory Strength between LS and Human Recall as a Function of Story

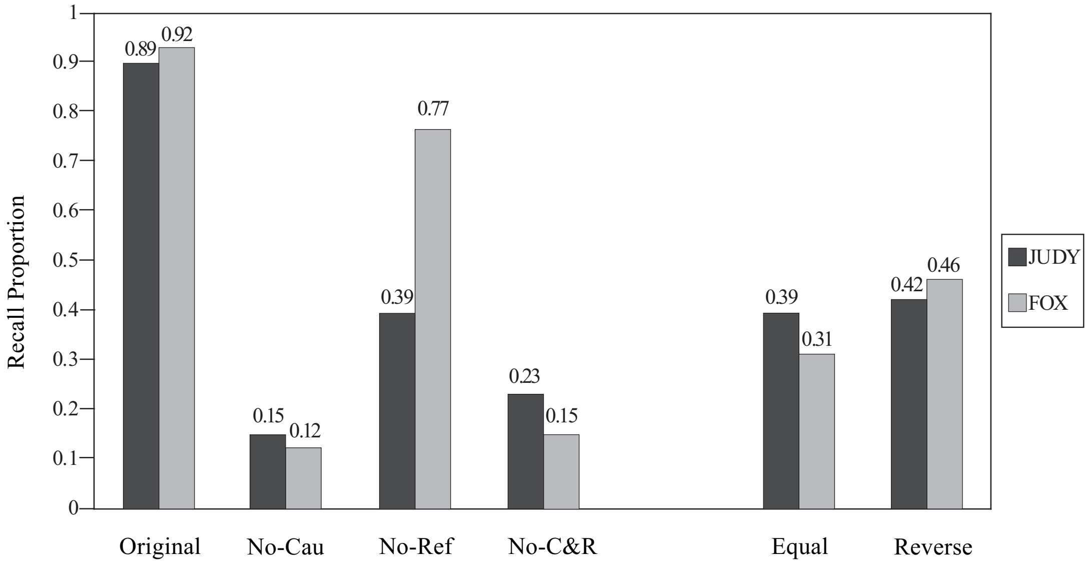

Figure 3. Proportion of Recall Predicted by Modified Versions of LS as a Function of Story. (No-Cau represents no causal connections. No-Ref represents no referential connection. No-C&R represents no causal and referential connections. Equal represents equal activation values for all sources. Reverse represents reverse activation values for different sources.)

故事體文章的記憶研究：景觀模式的運作機制

曾玉村

國立中正大學師資培育中心

理解的計算理論被視為是研究理解過程所發生的認知歷程與理解結束所形成的記憶表徵的一種有效的方式。本文旨在探討一個特定的理解的計算理論－景觀模式，試圖檢驗與建立其理論的效度。實驗1-a 企圖尋找一組合適的參數組合得以有效預測參與者對短篇故事體文章的回憶量，進一步將這一組的參數值套用於實驗1-b，並且將模擬的結果再與參與者對另外的短篇故事體文章的回憶量比對，發現景觀模式預測力仍然良好，足證這個模式的效力並不完成依賴參數值的選定，而是該模式本身就有一定的效力。實驗2再以模擬的方式檢驗景觀模式對各種激發來源的預設值是否合理，結果發現如果改變原先預設的激發值的話，模式的預測力大為降低，足證預設的激發值具有合理性，也支持該模式假定之因果推論及指稱詞推論卻實是理解過程必要之推論機制。本文也討論諸多本研究的限制與未來的研究方向。

關鍵詞：理解的模擬、景觀模式、文章記憶