

Comprehending Text: Achieving Coherence through a Connectionist Architecture

Samuel W.K. Chan

Department of Artificial Intelligence
School of Computer Science and Engineering
University of New South Wales
Sydney NSW 2052 AUSTRALIA

Abstract

Comprehension is determined by the content of memory. The verbalized concepts reflect the information that is in the focus of attention. Mental models of text must be formed in order to achieve coherence in language understanding. Constructing a mental model requires continual interaction between the text and the reader's linguistic, pragmatic, and world knowledge. In this article, a connectionist framework is proposed for the formation of mental models that occurs during comprehension. A unification and spreading activation model are used to simulate the processes. The system is tested using children's stories. Results attest the validity of the model. Thus, our design contributes to the formation of a durable and functional mental representation of the text information.

1. Introduction

Language comprehension is mediated by a complex set of processes that operate on the representations at three main different levels: lexical, syntactic and semantic. Early research on aspects of text representation tended to look at general questions about, for example, the availability of representations of surface form. More recently, comprehension is modeled as construction of mental models of text [1]. Several conclusions about mental representations of discourse have been reached. First, such representations are structurally similar to part of the world rather than to any linguistic structures [2]. "Distilled" meanings of the text and their interrelations are represented directly. This representations are built up as the text is read. Information not explicit in a text may be included in the representations of its content, particularly if it is required to establish links between parts of its content [3]. At any point in a text, the representation constructed up to that point is the context for the interpretation of the next sentence. In particular, it restricts the set of possible referents for anaphoric expressions, and allows such expressions to be interpreted correctly in context [4]. Given this agreement as to the theoretical importance of mental models, however, it is surprising that there is little agreement as to exactly what constitutes a mental model, and there is little research demonstrating the construction of a putative mental model while reading. In this article, our specific goal is to delineate a theoretical connectionist system that generates the cognitive representation for narrative comprehension. In particular, attention has focused on the factors of argument overlapping [5] and situational continuity

[6]. Researches in discourse processes have shown these two major factors facilitate readers to construct the mental models. Argument overlapping is analyzed in terms of the number of arguments overlapped. Propositions having the greatest degree of overlapping with other propositions are proved to be the most important. In addition, Gernsbacher [7] has advocated that readers construct mental structures while reading a text and they try to map any incoming information onto the evolving structure. Sentences that maintain a previously established time frame are more likely to be mapped onto developing structures, as are sentences that maintain a previously established location and sentences that are logical consequences of a previously mentioned action or event. For instance, it is relatively easy to construct a representation of a situation that involves a chronologically ordered chain of causally related events and action in one location. In such case, incoming situational information can be readily be integrated with the current mental model.

Applying connectionism in discourse processes has received a great deal of attention recently [8]. There are many extremely promising aspects of connectionist work that map quite well with the perspective developed within the spreading activation network. Spreading activation network is represented by a set of nodes fully connected to each other. It is a pattern categorization device inspired by neurophysiological considerations. The network has additionally been applied to fairly diverse range both in neural and psychological phenomena [9]. It provides a useful model of human learning and concept formation [10]. In this article, a brief discussion on how a sentence is distilled in a unification process is first presented. Second, a spreading activation network, which nodes are the propositions and the links are derived from the effects of argument overlapping and situational continuity, is formed and trained from the narratives. It is based on the theory that raw linguistic knowledge is said to undergo a reduction to essences and are encoded for storage and recall. The essences are often used to refer the residue of information that remains after a delay. Human inferences are the predisposition of reasoning to operate on traces that are as near as possible to the essences formed from past experience [11].

2. Overview of System Architecture

The whole system consists of several major constituents: the parser, logical and grounding tiers, working memory (WM), sentence memory and a long-term memory (LTM) as shown in Figure 1. The LTM has procedural component. The procedural LTM stores the causal hypotheses and a set of commonsense inference rules. Firstly, all the linguistic cues are used to retrieve knowledge from the LTM after parsing. Vectors of knowledge or rules are recalled and displayed on the logical tier. The logical tier represents the localist representations of concepts which are connected to the other relevant nodes. The grounding tier contains all the subsymbolic representations of concepts in the logical tier. The representations encode the knowledge associated with the concepts. Every concept in the logical tier is linked with its corresponding

subsymbolic representation in the grounding tier. The grounding tier roughly corresponds to reasoning at a subconceptual level.

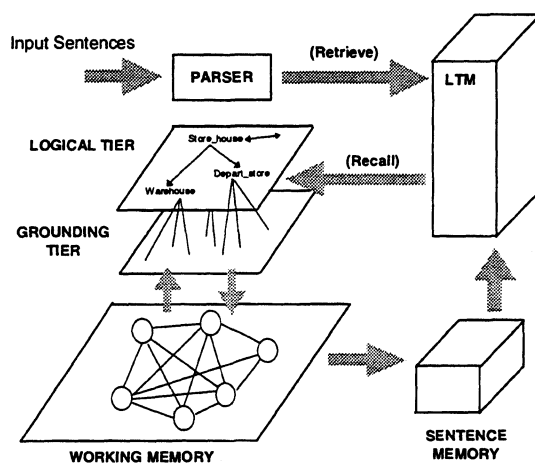


Figure 1: System Architecture

During comprehension, a unification process takes place in working memory in order to resolve the semantic ambiguity. After the process, the elements are contextually relevant and all contradictions have been eliminated. The working memory then contains the “distilled” items, associative and contextual information of the investigated sentence and is then transferred into the sentence memory.

In a separate strand, to call a sequence of sentences a “narrative” is to imply that the sentences display some kinds of mutual dependence; they are not occurring at random. These dependence may be lexical, logical, spatio-temporal or thematic. In order to represent these dependence in our analysis of narrative comprehension, in concert with all sentence memory in the narrative, a spreading activation network, W , with each node representing the distilled proposition after the unification processes, is constructed. The weights, W_{ij} , is relied on two major linguistic dependence (i) the argument overlapping (ii) the situational continuity, i.e. whenever the propositions are related, in the sense of the dependence, the weights between them are assigned a positive number. Thus, high activation in ones will produce high activation at the others. The network is then subject to a learning process, in which it will eventually reach a state of equilibration via the interactive activation and competition mechanisms. As a result, a stabilized network, a mental-like model, is obtained. Further discussions of the processes are as follows.

3. The Unification Process

In working memory, knowledge is represented as an associative net and retrieved from both tiers. Each element, or node, of the net has an associated activation level. The elements may represent words, phrases, propositions, grammatical structures or objects in the external world. Connections among elements have strength values. The strength values are calculated through a knowledge extraction process. The process needn't be very sophisticated. It is to

be just powerful enough so that the right elements are likely to be among those generated, even though irrelevant or even outright inappropriate will also be deduced. It is followed by a unification process. The process is used to integrate the meanings in the memory into a coherent whole. It is used to strengthen the contextually appropriate elements and inhibit unrelated and inappropriate ones, in which smart and complex deductions can be achieved.

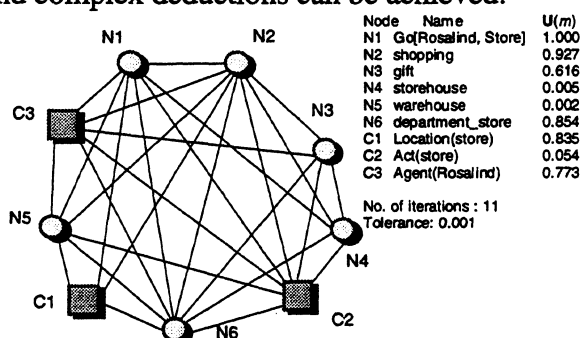


Figure 2: After reading a sentence, *Rosalind went to a store to buy a present*, a fully interconnected network with rough, or even outright contradictory nodes is formed and subjected to a unification process in working memory. The top right column shows the asymptotic activation after the process. It is apparent that incorrect meaning of *store*, *storehouse*, is deactivated.

This approach is to reduce the dimensionality of the stimulus so that a very complicated stimulus could act as if only a small number of independent elements are involved. In other words, this process can be used to exclude unwanted elements from the working memory, as an example shown in Figure 2. Some of the main properties of the unification process are summarized as follows:

- (i) Each working memory element represents a proposition or concept.
- (ii) Associated with each working memory element is a real number called its activation level, which represents the element's strength. Let a vector \mathbf{U} is modelled in N -dimensional space, then $\mathbf{U} = (u_1, u_2, \dots, u_N)$, where u_i is the activation level for the element i .
- (iii) Production firings direct the flow of activation from one working memory element, the source, to another working memory elements through the knowledge matrix \mathbf{K} . $\mathbf{U}(t+1) = \mathbf{T}[\mathbf{U}(t)\mathbf{K}]$ where $\mathbf{U}(t)$ is the activation vector at some discrete time t , \mathbf{T} is the vector normalizing operation.
- (iv) The process stops at iteration m if $|\mathbf{U}(m) - \mathbf{U}(m-1)| < t_u$ where t_u , the tolerance, is another preset threshold which is used to control the number of inhibited elements in the process.

Initially, $\mathbf{U}(1)$, representing the initial activation values of all concepts, is passed into the knowledge matrix, \mathbf{K} . Mathematically, unification process is a vector-matrix multiplication. Continued spreading by repeated vector multiplication leads to equilibration. The final activation vector, $\mathbf{U}(m)$, shows how strongly related items have strengthened each other, while unrelated or contradictory items have a negative or 0 activation value. The process continues

for the next sentence as another cycle in the narrative. The details of the unification process is omitted intentionally due to the limited scope of this article, however, the interested readers are referred to the author's publications [12-16].

4. Learning in Spreading Activation Network

Text processing is a sequential process. When a new sentence is comprehended, part of the previous sentences must also participate in the understanding. In Halliday's terminology [17], sentences are cohesive to the extent that they remain many expressions whose interpretation depends in some way on the interpretations of the prior expressions in the narrative. In order to examine the factor in comprehension, the previous analyzed sentences are maintained in the focus of attention when a new sentence is investigated in our unification process. Buffers are designated to carry the previous most activated propositions over into the current processing cycle, in the hope that they would serve as common bridging elements between the sentences. The change of the activation in each buffer indicates the degree of overlapping of each previous proposition to the current investigating proposition. Let $AO = (AO_{ij})$ be a square matrix representing the weights defined only by the effect of argument overlapping, or the cohesive links.

$$AO_{ij} = \exp(-\epsilon |\mu_{ij}|) \quad \text{where} \quad (1)$$

$$\mu_{ij} = \left(\frac{\text{change of activation of buffer } i}{\text{activation of buffer } i} \right) \Bigg|_{\text{Proposition } j}$$

Cohesive links go a long way towards explaining how the sentences of a narrative hang together, but they do not tell the whole story. A narrative plainly has to be cohesive as well as coherent, in that the concepts and relationships expressed should be relevant to each other, thus enabling readers to make plausible inference about the underlying meaning. Although causality, spatiality, and temporality are the intertwining links in the situational continuity when behavioral episodes are unfolded in narratives, we limit our scope in this article to the discussion of causal continuity, simply because it is ascertained that readers attempt to tie each event or fact encountered to the prior text or relevant background knowledge [18]. To capture the causal relationships, each of the subsequent statements is read in terms of whether it instantiates some expectation of the previous statements. Let $SC = (SC_{ij})$ be a square matrix representing network connections defined only by the effect of causal links. In our simulations, all pairs of processing units that are involved in the causal chain are given a mutually excitation connections,

$$SC_{ij} = \begin{cases} \xi > 0, & \text{if units } i, j \text{ are coherent related} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

An overall matrix, W , of the spreading activation network, as shown in Figure 3, is constructed by a combination of each independently linguistic dependence discussed above.

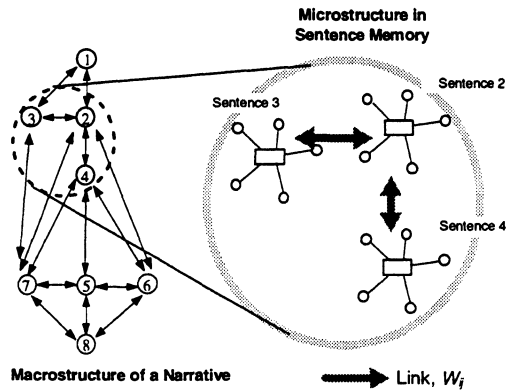


Figure 3: The distilled propositions are linked with each other in sentence memory in form of a spreading activation network, W , where the links, W_{ij} , are determined by the factors of argument overlapping and situational continuity.

Each entry W_{ij} in W is a numerical value that represents the connection strength between the propositions. Further, our model takes the simplest conceivable step of generalization by substituting a pair of oppositely-directed links of equal weight. Each of connections carries the meaning, “is related to” without discriminating the roles of cause or effect so that W is a symmetric matrix; that is $W_{ij} = W_{ji}$ such that

$$W_{ij} = \begin{cases} \rho(\omega_1 AO_{ij} + \omega_2 SC_{ij}), & \text{if } AO_{ij} \text{ or } SC_{ij} \neq 0 \\ -\eta, & \text{or otherwise} \end{cases} \quad (3)$$

In (3), it is apparent that positive excitatory links, with values $\rho(\omega_1 AO_{ij} + \omega_2 SC_{ij})$, are between all the linguistic related nodes and negative inhibitory links, $-\eta$, are amongst the unrelated ones. ω_1 , ω_2 are constants specifying the relative importance of each effect and ρ is a coefficient that determines the rate of convergence of the algorithm. Since W is symmetric, this ensures that the spreading activation network will converge to a stable state. Whenever the matrix W has been constructed, one can start the learning process to encapsulate the mental-like model from the network, W . The model extracted via the interactive activation and competition mechanisms. The algorithm of the encapsulation of the model is summarized as follow:

Step 1: Construct the overall matrix, W , as shown in above. Initialize a state vector for the propositions,

$A(0) = (a_1(0), a_2(0), \dots, a_M(0))$ where $a_i(0)$ is the initial activation of the proposition i and M is the number of the propositions in the narrative.

Step 2: At each discrete time, t , activations spread among the nodes of propositions and are updated by the following function:

$$A(t+1) = \beta A(0) + \gamma A(t) + \alpha W A(t)$$

Step 3: Vector $A(t+1)$ is normalized according to the saturation and habituation functions

$$A(t+1) = [SAT(A(t+1))] \sigma(t) \text{ where}$$

$$SAT(x) = \begin{cases} 1, & 1 < x \\ x, & -1 \leq x \leq 1 \\ -1, & x < -1 \end{cases}$$

Step 4: Reinforcement of W_{ij} , is given by the modified Hebbian learning, where

$$\delta W_{ij}(t) = \phi\{[a_i(t+1) - a_i(t)]a_j(t) + [a_j(t+1) - a_j(t)]a_i(t)\} \text{ where } \phi \text{ is the learning rate.}$$

Step 5: Apply steps 2 to 4 iteratively for a number of iterations

Step 6: The final link strength between each pair of node i, j in model is

$$W_{ij}^* = \Psi(W_{ij}) \text{ where}$$

$$\Psi(x) = \frac{1}{1 + e^{-\lambda(x - \bar{x})}} \text{ } \lambda \text{ is the gain, } \theta \text{ is the event linkage constant, and } \bar{x} \text{ is the mean of } x.$$

The algorithm is the modified version of Brain-state-in-a-Box model (BSB) with M nodes which are highly interconnected and feedback upon themselves. Information in this system is represented by an M -dimensional vector. The system operates by accepting a pattern of neural activity and amplifying that neural activation pattern through the feedback loop. In step 2, the system receives a constant input, $\beta A(0)$, in which the initial activation of each proposition is constantly present. The second term, $\gamma A(t)$, causes the current state to decay slightly. This term has the qualitative effect of causing error to eventually decay to zero as long as γ is less than 1. The third term, $\alpha WA(t)$, passes the current state through the matrix and adds more information reconstructed from the cross connections. The saturation function $SAT(x)$ is the nonlinearity in the system in step 3. This confines the BSB state to the hypercube $[-1, 1]^M$. When released from the initial state, the BSB converges to one of the stable vertices of the hypercube. In addition, $\sigma(t)$ is introduced as a habituation variable into the system. It provides a mechanism to get the brain state out of a corner it has gotten into. Once a node has reached its maximum firing rate, this process becomes effective over a fixed period of time and lowers the input sensitivity of the node, thus causing a decrement in the components of the saturated subvector, as a consequence, which may leave the corner.

The formation and the reinforcement of the weights are represented in step 4. The changes in strength of the weights are through modified Hebbian learning. It is apparent that the strongest connections will tend to form between pairs of nodes that maintain high levels of activation for a prolonged period of time. Finally, the function Ψ , in step 6, serves as a link transfer function. It is a sigmoid function that enhances the effects of links by the gain λ with strength above θ , the event linkage constant multiply with the mean W_{ij} . It controls the amount of links appearing in the contexts. This final stage of the encapsulation is to screen the links and reduce background clutter in the emerging representation. The selection process creates a threshold below which a link is unlikely to be emitted as part of the output in the final mental-like model. It eliminates the weaker connections which result from the formation of weak association.

5. Simulations and Experimental Results

The prototype of the system is developed in C and implemented on a DEC 5000/20 under the UNIX environment. Simulation experiments are conducted. We restrict ourselves, in this article, to stories describing sequences of events which follow one another in approximately linear temporal sequences. However, the ideas suggested here are certainly not only applicable to this restricted class of stories. In our experiments, four children's stories by Stein and Clenn [19] are analyzed in a series of simulations. Tables 1 & 2 show one such story, *Tiger's Whisker* and the corresponding propositions so formed.

Once there was a woman who needed a tiger's whisker. She was afraid of tigers but she needed a whisker to make a medicine for her husband who had gotten very sick. She thought and thought about how to get a tiger's whisker. She decide to use a trick. She knew that tigers loved food and music. She thought that if she brought food to a lonely tiger and played soft music the tiger would be nice to her and she could get the whisker. She went to a tiger's cave where a lonely tiger lived. She put a bowl of food in front of the opening to the cave. Then she sang soft music. The tiger come out and ate the food. He then walked over to the lady and thanked her for the delicious food and lovely music. The lady then cut off one of his whiskers and ran down the hill very quickly. The tiger felt lonely and sad again.

Table 1: The *Tiger's Whisker* story

| | |
|-------------------------------------------------|----------------------------------------|
| 1. Exist(Woman) | 12. Think(She, Get(Whisker)) |
| 2. Afraid(She, Tigers) | 13. Go(She, Tiger's Cave) |
| 3. Need(She, Whisker) | 14. Live(Tiger, Tiger's Cave) |
| 4. Desire(She, Make(She, Medicine, Husband)) | 15. Put(She, Food, Front of Cave) |
| 5. Sick(Husband) | 16. Sing(She, Music) |
| 6. Think(She, How_to_get(She, Tiger's Whisker)) | 17. Go(Tiger, Front of Cave) |
| 7. Desire_to_use(She, Trick) | 18. Eat(Tiger, Food) |
| 8. Know(She, Love(Tiger, Food and Music)) | 19. Walk(tiger, Lady) |
| 9. Think(She, Bring(She, Food, Lonely(Tiger))) | 20. Thanks(Tiger, Her, Food and Music) |
| 10. Think(She, Play(She, Music, Lonely(Tiger))) | 21. Cut(Lady, Tiger's Whisker) |
| 11. Think(She, Nice(Tiger)) | 22. Run(Lady, Down Hill) |
| | 23. Sad(Tiger) |

Table 2: The Propositions of the *Tiger's Whisker* story

Figure 4 shows the activation changes of some previous propositions while the 6th proposition, Think(She, How_to_get(She, Tiger's Whisker)), is under analyzed. Clearly, the activation of the 5th proposition decays gradually to zero due to the nonoverlapping arguments. In contrast, activations of the 3rd & 4th propositions remain relatively high during the analysis. In our simulations, ϵ , in equation (1), is chosen to be 0.75. The final matrix, W , of the spread activation network is formed and is subject to the learning algorithm as discussed in section 4. Re-arranged patterns of activation of the propositions are shown in Figure 5. The three-dimensional mesh plot is interpreted by noting that the height of each point in the mesh plot corresponds to the activation of a time-dependent feature unit in the system. All points lying on the same horizontal line correspond to the same proposition (e.g. 8. Know (She, Love (Tiger, Food and Music))) at different points in time. All points lying on the same vertical line correspond to the values of all features associated with a situation at a particular instant of time where

time is ordered from the left-hand side of the graph to the right-hand side. In Figure 5, the activation of the 23 propositions in the networks are indicated by the height above or below the null activation level. It is apparent, after 14 iterations, the network reaches a state of equilibration eventually. All propositions become saturated with activations $\{-1, 1\}$ and the mental-like model is thus formed.

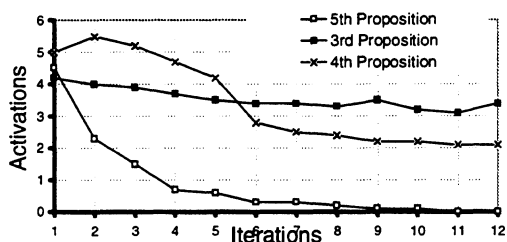


Figure 4: The activations change in the 3rd, the 4th and the 5th propositions while the 6th proposition is under analysis in the story. The words, *she*, *tiger's whisker*, in the sixth proposition are the overlapped arguments which makes the activations in the corresponding propositions remain high.

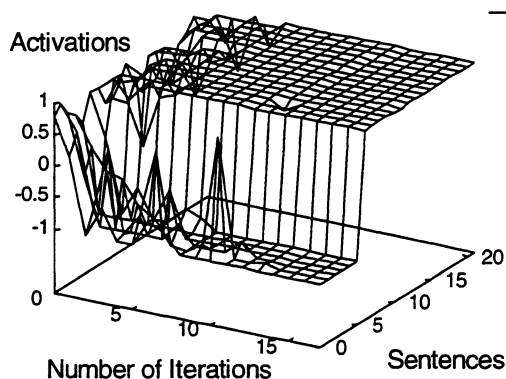


Figure 5: Change of the activation of the propositions in the story of *Tiger's Whisker*, during learning

6. Conclusions

In this article, we have discussed the issues in linguistic information that will effect the comprehension. We are able to identify the general “thread” of the discourse as well as the way that individual sentences fit together to achieve the comprehension in narrative understanding, in particular, in the formation of the deep level of the mental representation of discourse. The present study explores the cognitive representation of narrative prose. A distinctive property of the proposed architecture is that it incorporates the inferences that readers generate on the basis of their knowledge about the world and language. An attempt is also made in the article to generate the mental model. Our construction process can be interpreted as a particular mental transformation of a given set of initial sentences and concepts, based on the linguistic relations. Experiments show the representation of texts as spreading activation networks captures important aspects of the memory representation of the text and it may suggest one of the formation of mental models in humans.

References

- [1] Johnson-Laird, P.N. (1983). *Mental Models*. Cambridge, MA: Harvard University Press.
- [2] Garnham, A. (1981). Mental models as representations of text. *Memory & Cognition*, 9, 560-565.
- [3] Garnham, A., Oakhill, J., & Carreiras, M. (1995). Representations and processes in the interpretation of Pronouns: New evidence from Spanish and French. *Journal of Memory & Language*, 34, 1, 41-62.

- [4] Rinck, M., & Bower, G.H. (1995). Anaphora resolution and the focus of attention in situation models. *Journal of Memory & Language*, **34**, 1, 110-131.
- [5] van Dijk, T.A., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic.
- [6] Zwaan, R.A., Magliano, J.P., & Graesser, A.C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 2, 386-397.
- [7] Gernsbacher, M.A. (1990). *Language Comprehension as Structure Building*. Hillsdale, NJ: Erlbaum.
- [8] Miiikkulainen, R., & Dyer, M.G. (1991). Natural language processing with modular PDP networks and distributed lexicon. *Cognitive Science*, **15**, 343-399.
- [9] Anderson, J.A., & Murphy, G.L. (1986). Psychological concepts in a parallel system. *Physica D*, **22**, 318-336.
- [10] McClelland, J., & Rumelhart, D.E. (1985). A distributed model of human learning and memory. In J. McClelland & D. Rumelhart (Eds.), *Parallel Distributed Processing: Explorations in microstructure of cognition*, Vol. 2, MIT Press, 170-215.
- [11] Ellis, H.C., & Hunt, R.R. (1989). *Fundamental of Human Memory and Cognition*. Dubuque: Brown.
- [12] Chan, S.W.K., & Franklin, J. (1994a). A neural network model for acquisition of semantic structures. *Proceedings of IEEE International Symposium of Speech, Image Processing & Neural Networks*, Hong Kong, 221-224.
- [13] Chan, S.W.K., & Franklin, J. (1994b). An epistemological model in semantic cognitive map. *Proceedings of International Conference on Neural Information Processing, ICONIP'94*, Seoul, Korea, 1073-1078.
- [14] Chan, S.W.K., & Franklin, J. (1994c). Symbolic connectionism in tiers: A strategy of discourse comprehension. *Proceedings of Seventh Australian Joint Conference on Artificial Intelligence*, Australia, 434-441.
- [15] Chan, S.W.K., & Franklin, J. (1995a). Inferences in natural language understanding. *Proceedings of the IEEE Fourth International Conference on Fuzzy Systems, FUZZ/IEEE'95*, Yokohama, Japan.
- [16] Chan, S.W.K., & Franklin, J. (1995b). A neurosymbolic integrated model for semantic ambiguity resolution. *Proceedings of the IEEE International Conference on Neural Networks*, Perth, Australia.
- [17] Halliday, M.A.K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- [18] Trabasso, T., & Sperry, L.L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language*, **24**, 595-611.
- [19] Stein, N.L., & Clenn, C.G. (1979). An analysis of story comprehension in elementary school children. In R.O. Freedle (Ed.), *New Directions in Discourse Processing*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 53-120.