A Connectionist Implementation of Passive Elaborative Inferencing

M. Iza[†], N.S. Park[‡]

[†]Dpto. de Lógica y Filosofía do la Ciencia
Universidad del País Vasco/Euskal Herriko Unibersitatea, Spain

[‡]Dept. of Artificial Intelligence, University of Edinburgh, Scotland

Abstract

O'Brien et al. (1988) reported that readers generated elaborative inferences only when a text contained characteristics (a strong biasing context or a demand sentence) that made it easy to predict the specific inference that a reader would draw and, virtually eliminated the possibility of the inference being disconfirmed. Garrod et al. (1990), however, offered two refinements to the conclusions. First, the two text characteristics manipulated may have produced different types of elaborative inferencing: a biasing context results in a passive form of elaborative inferencing, involving setting up a context of interpretation, whereas the presence of a demand sentence invites the reader to actively predict a subsequent expression. Secondly, clear evidence for either type of inference will be apparent only with truly anaphoric materials.

This paper describes how a passive form of elaborative inferencing, reported by Garrod et al, may be implemented in a connectionist manner. We used the connectionist model proposed by Shastri & Ajjanagadde (1993) to represent a text containing characteristics in the form of a network. And we have recently adopted meta-predicates to represent relationships between contexts and anaphors. These meta-predicates are encoded into the network and used to control the flow of inference according to the relationships specified by them.

1 Introduction

During reading comprehension, human behaviour data show that representation of conceptual information and the cognitive processes that interact with it are such that not only are relevant facts automatically retrieved, but certain kinds of inferences also exhibit extreme efficiency. Analysing the resulting memory trace, we find that human agents perform a broad range of inferences, according to explicit stated information and with information based on world knowledge, in an automatic way.

These types of inferences has been classified in two general classes: necessary and elaborative. In the first class, readers perform necessary inferences such as those needed to maintain referential coherence (e.g., McKoon & Ratcliff, 1986; O'Brien, Duffy & Myers, 1986) or to establish causal relations (e.g., Myers, Shinjo & Duffy, 1987). And, in the second class, readers draw elaborative inferences to predict upcoming consequences or information (e.g., McKoon & Ratcliff, 1986; O'Brien, Shank, Myers & Rayner, 1988; Garrod, O'Brien, Morris & Rayner, 1990). For instance, perhaps as soon as comprehenders read that "The actress fell from the 14th floor", they predict that the actress died. If so, they would be generating what is called a predictive inference. However, the data show that this inference is only partially encoded into the mental representation of a text (McKoon & Ratcliff, 1986).

Similarly, during anaphora processing, its interpretation can depend upon some arbitrary part of general knowledge not easily accessible. So, the anaphora needs inference processes, based on world knowledge or, on context, that have to be performed quickly. As we shall see later, some of these inference processes can be forward inferencing processes that help in the interpretation of a subsequent anaphora.

In the remaining part of the paper, we describe how a particular kind of forward inferencing could be implemented in a connectionist manner. We explore this on a connectionist model proposed by Shastri & Ajjanagadde (1991, 1993). Their model can represent knowledge in the form of 'facts' and 'rules' in a connectionist manner and determine whether a given 'question' can be answered or not by facts and rules encoded in the system. The model can also represent constants, types (categories), the instance-of relations between constants and types, and the super/sub-concept relations between types in the form of IS-A hierarchy.

This model, however, only controls the adequacy of argument-fillers and leaves untouched the activation control issues. In implementing forward inferencing, there exists the danger that there are too many available inferences. People perform some forward inferencing on the guess that the same patterns occur. But they do not carry them vary far because the certainty of the inferences quickly falls below some threshold of plausibility. So, we have to address the activation control problem with Shastri & Ajjanagadde's model.

As we shall see in Section 3, meta-predicates are used to control the flow of activation during inference. These predicates represent the relationship between contexts and anaphors and are encoded into the network together with rules and facts.

2 Psychological evidence for forward inferencing

A crucial question for theories of inference during reading concerns how to characterise the inferences people make and the circumstances under which they make them. As O'Brien et al. (1988) argued there are two majors sources of difficulty in the study of forward inferences. First, as they are not necessary for comprehension, it is difficult to predict exactly if and when a reader will draw and inference or to know exactly what that inference might be. Secondly, readers may generate elaborative inferences only in limited situations, choosing to delay any inferential process until it becomes necessary.

In general, investigations of the inference processes that occur during reading have usually been designed to demonstrate that some particular type of information is or is not inferred. Recently, McKoon & Ratcliff (1986) have proposed an alternative framework for research in this area. The principal notion is that inferences are not necessarily encoded into the memory representation of a text in an all-or-none fashion, but instead can be encoded in varying degrees. The degree of encoding can vary from not encoded at all to partially encoded to exemplars encoded. For McKoon & Ratcliff, a partially encoded inference is a set of features of meaning that does not completely instantiate the inference though attributes of meaning might equally well be encoded as propositions.

Examples of inferences with different degree of encoding have been provided in several experiments that examined the degree of encoding of inferences about predictable events and contextually defined categories. For example, McKoon & Ratcliff (1989) show that inferences about the most typical exemplars of such categories are encoded to a high level during reading and that the content of the inference is made up of information relating properties of the most typical exemplar to textual information.

In this way, McKoon & Ratcliff (1989) modify their general framework of minimal inference processing: If a specific inference is provided by easily available general knowledge from long-term memory, then it will be constructed even if it is not required for coherence. But this modification applies only for the most easily available information; other inferences such as those about predictable events, about default values in schema representations, and about the instruments of verbs, are usually encoded partially or not at all.

For instance, in Sanford & Garrod's (1981) model it is almost misleading to say that given "Jane unlocked the door" the inference "She used a key" is drawn. For Sanford (1990) rather such 'knowledge' becomes 'accessible'.

On the other hand, O'Brien et al. (1988) reported that readers generated elaborative inferences only when a text contained characteristics that made it easy to predict the specific inference that a reader would draw and, virtually eliminated the possibility of the inference being disconfirmed. This occurred only when there was a strong biasing context prior to the point at which the inference was expected (see text 1 in Table 1) or when the text invited an inference by means of a demand sentence (see text 2 in Table 1). As soon as these characteristics were relaxed, evidence for the establishment of the inference disappeared.

However, Garrod et al. (1990) reanalyzed these data and offered two refinements to O'Brien et al.'s conclusions. First, the two text characteristics O'Brien et al. manipulated (a strong biasing context or a demand sentence) may have produced different types of elaborative inferencing. Garrod et al. argue that a biasing context results in a passive form of elaborative inferencing, involving setting up a context of interpretation, whereas the presence of a demand sentence invites the reader to

Text 1 All the mugger wanted was to steal the woman's money. But when she screamed, he [stabbed] [assaulted] her with his (knife/weapon) in an attempt to quiet her. He looked to see if anyone had seen him. He threw the *knife* into the bushes, took her money, and ran away.
Text 2 Joan was delighted when Jim gave her a ring with a (diamond/large stone) in it. [He had asked her to marry him, and now they were officially engaged.] [He often bought her expensive and unusual gifts, and this was no exception.] She went to show her father. *He asked what kind of gem it was.* She excitedly told him that it was a *diamond* from her boyfriend.
Note: The high-context version was created by including the text in the first set of brackets; the low-context version was created by using the text in the second set of brackets. The antecedent is in parentheses, with the explicit and implicit versions to the left and right of the slash, respectively. The target anaphora is in italics, and the demand sentence in the text 2 is in italics.

Table 1: Example passages used in O'Brien, Shank, Myers, and Rayner (1988)

actively predict a subsequent expression. Secondly, clear evidence for either type of inference will be apparent only with truly anaphoric materials.

In this way, Garrod et al. (1990) argue that, in the previous case of 'the actress', with sufficiently constraining context many more of the semantic features defining 'dead' would have been encoded, leading to the activation of a predicted inference. So, they conclude that it is clear that the conditions under which evidence for elaborative inferences will be evident is limited. One such condition is a highly restricting context.

Finally, inferences of all types (including elaborative activity) will be more frequent with main characters than secondary characters. This amounts to mapping into background knowledge which serves the perspective of the main character. A knowledge-based account of understanding with this kind of restriction, or preference, would be recruiting single-perspective knowledge, which we would normally acquire as a result of our experience. Also, this raises the interesting question of when and how a new entity is considered more topical than the previous topic entity.

3 An implementation of passive elaborative inferencing

As discussed in Section 2, a certain result of elaborative inference is affected by the relationship between a context and an anaphor. In the text 1 of Table 1, for example, the recognition time of the target anaphor, the *knife*, depends on a context used before. In the case of the restricting context sentence ("He stabbed her with his weapon") or of less restricting context ("He assaulted her with his weapon"), there was a reliable interaction between context (restricting vs. nonrestricting) and lexical specification (explicit vs. implicit introduction of the antecedent). Readers encountering strong restricting context seemed to infer a more specific referent for the antecedent than would be licensed by its initial description. However, this interaction disappeared with the relaxation of the restriction of context. Thus the relationship between context and an anaphora plays an important role in understanding written text.

This section firstly describes how the connectionist model proposed by Shastri & Ajjanagadde, as a reflexive reasoning model, could be used for forward inferencing. Secondly, we show how this model can be extended for the elaborative inference which takes account the relationship between a context and an anaphora during inference.

A connectionist approach based on reflexive reasoning

Shastri & Ajjanagadde (1993) have proposed a partial connectionist model that can represent a large body of systematic knowledge and can perform some range of inferences (forward and backward reasoning) on structured knowledge in the LTM. Their model has shown in detail a technique how 'neuron-like' elements can be used to encode facts and rules involving n-ary predicates with variables.

A key part of their mechanism to solve the dynamic variable binding problem is to match rhythmic patterns of activity using temporal synchrony within the network. The temporal features on this temporal synchrony solution lies in a separation of a period of oscillation cycle into several phases, and in the use of phase sensitive nodes. On the basis of these properties, dynamic variable binding is represented by argument nodes and their constant, sharing the same phase in any particular period of oscillation during a reasoning process.

Consequently, their system is able to represent composite structures in a dynamic manner and systematically to propagate them so as to generate other composite structure. In this way, they propose a neurally plausible solution to the variable binding problem and describe a system that can represent a large body of systematic knowledge in the form of 'facts' and 'rules' involving n-ary predicates. Their system can also represent constants, types (categories), the instance-of relations between constants and types, and the super/sub-concept relations between types using an *IS-A* hierarchy. Rules and is-a hierarchy, extracted from Text 1 in Table 1, are defined in their system as follows:

$$\forall X, Y, Z : weapon \quad stab(X, Y, Z) \rightarrow use(X, Z)$$
$$\forall X, Y, Z : weapon \quad assault(X, Y, Z) \rightarrow use(X, Z)$$
is-a(gun, weapon)
is-a(knife, weapon)

In order to encode these rules and is-a hierarchy, Shastri & Ajjanagadde's system builds a directed acyclic inferential graph where each node in the graph is a processing element and each connection is well-connected (hardwired) link as shown in Figure 1.

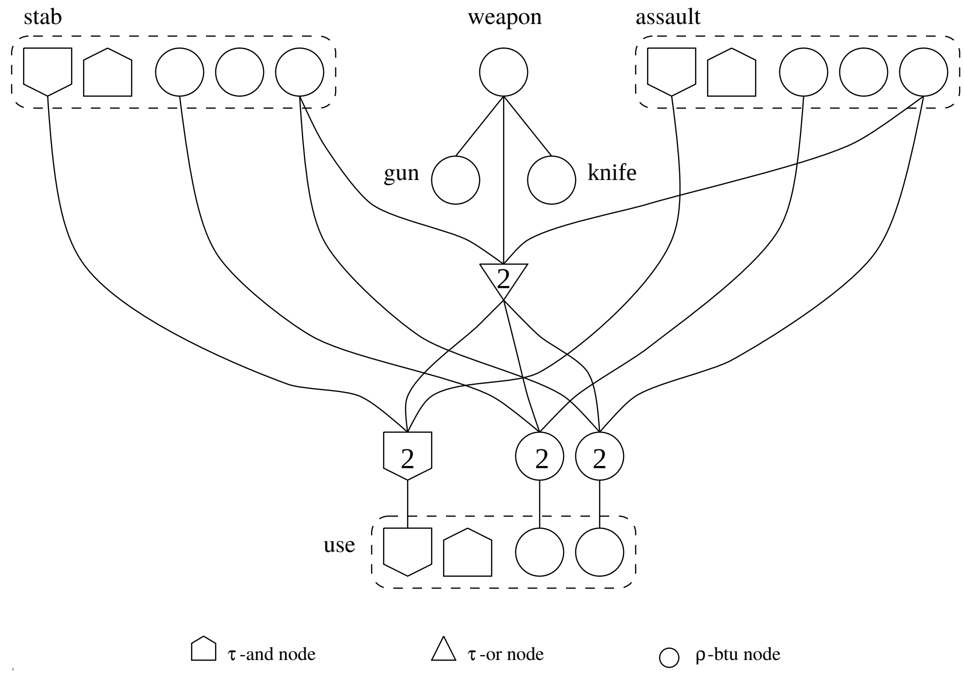

Figure 1: The encoding of the rules and is-a hierarchies

For more expressive power, the *IS-A* hierarchy, in their system, is used not only for representing hierarchical knowledge but also for imposing type restrictions on variables occurring in rules during forward inference. This interaction between rule-based reasoner and is-a hierarchy allows the system to encode context dependent rules that are sensitive to the types of the argument-fillers involved in particular situations. In the first rule, for example, the typed variable, *Z:weapon*, constraints the variable Z to be *weapon* type of fillers such as *gun* or $knife$.

Once world knowledge is encoded into the network, reflexive reasoning on the network starts by activating initial predicate nodes. After initial predicate nodes become active, this activation propagates all over the network through links between predicates. An activation of a certain predicate during inference represents part of a result of inference.

One of the advantages of their system is that knowledge can be encoded in a parallel way. Encoded rules and facts can be fired in parallel if they are connected to the same predicate nodes when the nodes become active. Thus the time taken for each inference depends on the length of a chain of inference and does not depend on the number of rules and facts encoded in the system.

However, there must exist constraints on the class of reasoning that may be performed in a reflexive manner. Not surprisingly, cognitive agents can perform only a limited class of inferences with extreme efficiency. Nevertheless, the proposed system only controls the argument-filler's adequacy, but leaves unexplained the activation control issue, that is to say, the question about why we perform just some inferences, and not others, in a given situation.

For example, forward inference between $stab$ and use in the network does not take into account the relationship between $stab$ predicate and the value of its third argument at all. In other words, if the query, $stab(mugger, woman, weapon)$ is posed to the network, their system equally activates gun and $knife$ nodes during inference and the system cannot differentiate whether the mugger used gun or $knife$. This is against what cognitive agents might get from the inference. Cognitive agents would have an image of a $knife$ as a $weapon$ rather than gun when they receive the same query. This is why we may predict that the mugger must have used $knife$ to stab the woman. The same result can be obtained when we pose $assault(mugger, woman, weapon)$ to the network. Again, in this case, although we expect the activation of only gun node in relation with the $assult$ predicate, but both $knife$ and gun nodes become active.

Controlling inference using meta-information

Representing passive elaborative inferencing in Shastri & Ajjanagadde's system requires one way of controlling activation flow among constant nodes in an is-a hierarchy. As we have addressed, this control of inference should be made based on the contexts and anaphors represented in the network. Thus, what we need is the way of representing the relations between contexts and anaphors and the way of encoding them into the network.

One possible way of representing the relations could be introducing meta-predicates. Here, we adopt the meta-predicate, *employed*, to represent such relations. If we use a predicate node to present a context and a constant node to represent an anaphor, the relations between contexts and anaphors can be represented using the definition of meta-predicates:

employed(type:constant,pred_name)

where $constant$ is a member of $type$ and has close relation with $pred\_name$.

This following illustrates how these meta-predicates are used to represent the relations in cooperation with existing rules and is-a hierarchy:

$$
\begin{array}{l}
\forall X, Y, Z : weapon \quad stab(X, Y, Z) \rightarrow use(X, Z) \\
\forall X, Y, Z : weapon \quad assault(X, Y, Z) \rightarrow use(X, Z) \\
is\text{-}a(gun, weapon) \\
is\text{-}a(knife, weapon) \\
employed(weapon : gun, assault) \\
employed(weapon : knife, stab)
\end{array}
$$

When we encode meta-predicates together with existing rules and is-a hierarchy, we need to encode them in such a way that the flow of inference changes based on a relation between a predicate name (a context) and a constant node (an anaphor) during inference.

One way of controlling inference in Shastri & Ajjanagadde's system would be by adopting the amount of activation and controlling it to distinguish a dominant node (whose activation value is higher than others) from other nodes. By this way, we can make a node which represent a certain anaphora becomes dominant when a predicate node which has a close relation with it becomes active. This way, however, requires many changes in the original connectionist model since it is designed to handle binary inputs for the binary outputs.

An alternative way may be using inhibitory links to control the flow of inference. This way encodes meta-predicates in the form of inhibitory links on top of the existing network. During inference, these links make some constant nodes which belong to a certain type of is-a hierarchy active and leave others inactive according to the relations specified by meta-predicates. In this way extra inhibitory links based on meta-predicates controls the flow of inference according to the relationship between a context and an anaphor. Figure 2 shows how meta-predicates can be encoded into the network in the form of extra inhibitory links.

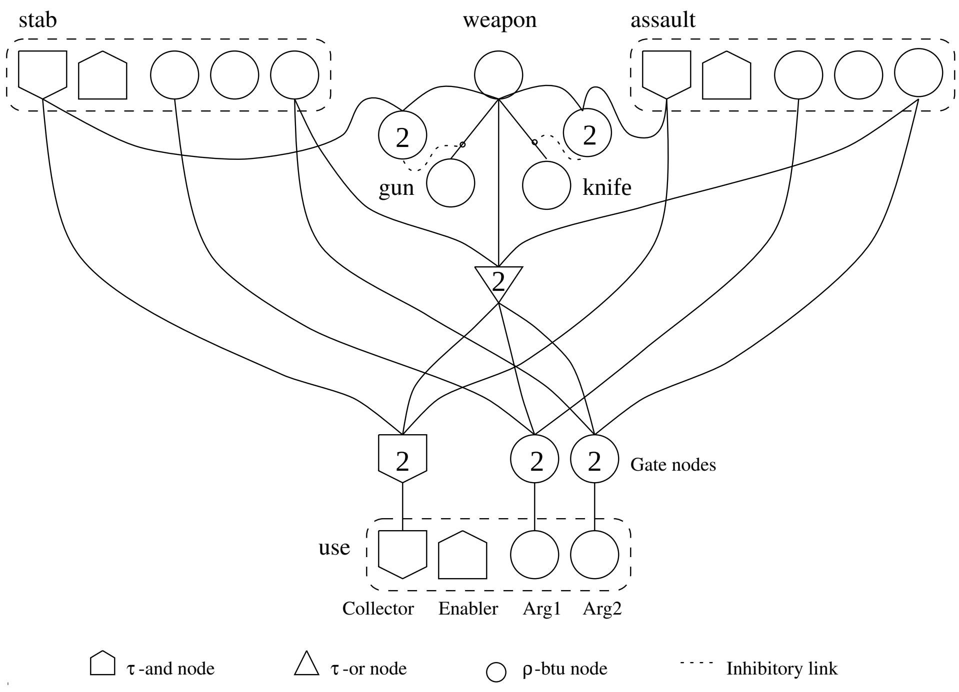

Figure 2: Network representation of rules and is-a hierarchies with meta-information

When the *stab* predicate and its third argument (type of weapon) become active during inference, the activation is used to project an inhibitory signal to the link between *weapon* and *gun*. Consequently, the *stab* context leads to activation of *knife* node but not of *gun* node.

Figure 3 shows phasic representation of inference when two queries are posed to the system. To the given query, $stab(mugger, woman, weapon)$, the *knife* node in the is-a hierarchy becomes active in the same phase *weapon* node, whereas the *gun* node becomes active to the query, $assault(mugger, woman, weapon)$.

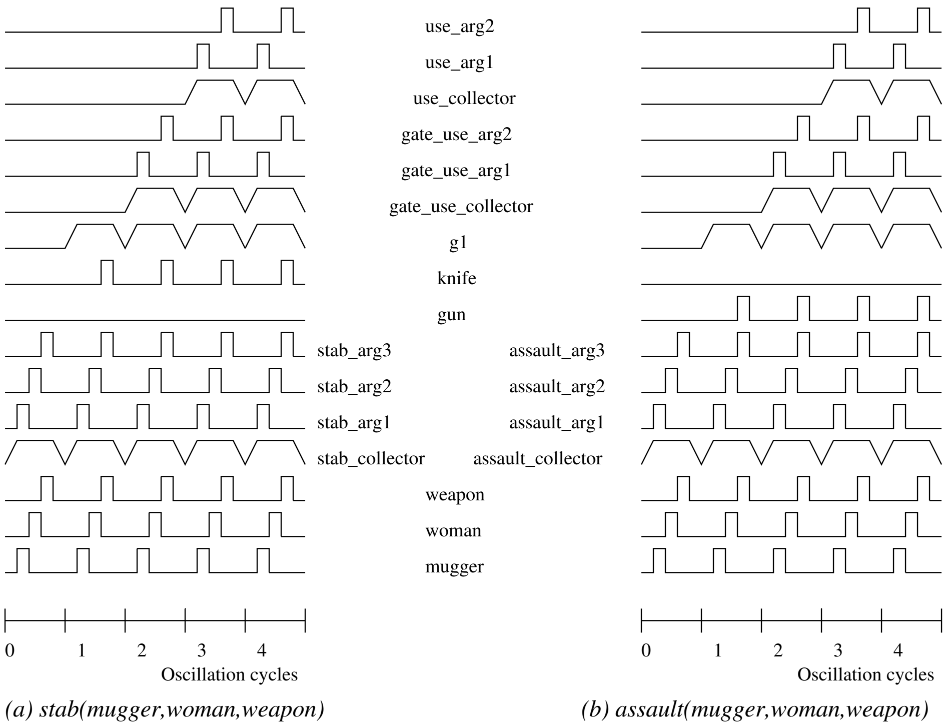

Figure 3: The phase representation of the two query processing

4 Discussion

Traditionally, a text has been thought of as a set of connected propositions, with inferences either filling in gaps or elaborating the structure. Either way, it makes sense to say that all inferences will be predicate-argument form. However, it can be argued that some of the work required to produce an inference might get done without the complete inference being calculated.

It is assumed that inferential activity may depend upon a continuous checking of an indeterminate number of input signals against norms for each signal type (syntactic, semantic, textual, conceptual and probabilistic), where the norms might be determined locally by the text or globally by more general knowledge based on expectations (see, Stevenson, 1993). This way of thinking about inference generation presents insurmountable problems for conventional AI modelling and seems to be naturally explicable within the connectionist frameworks. For example, norm violation is one thing that pattern-matching mechanisms are capable of dealing with (also matches themselves). During comprehension, reference resolution seems to be reliant on such processes (Sanford, 1990).

As we have seen above, a different representation of each argument and predicate together with an explicit encoding of inferential chain, between argument predicates, seems essential if a system is wanted to be applied to a large number of dynamic binding that results from these rule applications.

However, although the system proposed by Shastri and Ajjanagadde could be extended to treat some open problems in order to account for the sort of inferences performed in a given situation, its implementation for forward inferencing seems to be more complicated if we take into account the above experimental results.

In this sense, we have adopted meta-predicates in order to control the flow of inference according to the relationships specified by them. Similarly, Noordman & Vonk (1992) have shown that the reader's knowledge of the world is an important factor in controlling inferences. In their work, the role of the reader's knowledge with respect to the information to be inferred is investigated by varying the materials in terms of their familiarity to the reader (novice vs. expert).

In general, during text processing, readers must have some domain's knowledge to be capable of building event representations. For example, knowledge about properties of entities that produce the event, the kind of relationships in which these entities can enter, and how this relationship can be represented. With highly familiar domains, this knowledge can remain easily available, so providing the possibility of building mental models that help both automatic inference performance and anaphoric reference interpretation.

Further work

In the case of elaborative inferencing as a passive process, we have seen that meta-predicates can be used to control the flow of activation within the network. However, we have to take into account other kinds of elaborative inferences proposed by the literature [see for example Sanford 1990, Noordman & Vonk 1992] as well.

One example can be when we see elaborative inferencing as an active process, such as occurred in text 2 of Table 1. When the text invited an inference by means of a demand sentence, the reader actively predicted a subsequent expression. Currently, we are studying how this kind of process could be implemented in the proposed system. A possibility could be to consider the demand sentence as a question (backward reasoning) to the network. In this way, we have to address the question of how to control the interaction between backward and forward reasoning.

Finally, it seems also necessary to take into account the role played by the different characters in a text. As it has been mentioned above, elaborative activity seems to be more through main characters than to secondary characters. In the proposed system, we need both to differentiate the degree of activation of these characters (and its role in controlling inferencing), and, to deal with the question of when and how this difference of activation may be modified. For instance, it is possible that a given elaborative inference could produce a change in the activation of the characters (i.e., a focus shift).

References

[Garrod et al. 1990] S. Garrod, E. J. O'Brien, R. K. Morris & K. Rayner. Elaborative inferencing as an active or passive process. *Journal of Experimental Psychology: Learning, Memory, and cognition*, 16(2):250–257, 1990.

[McKoon & Ratcliff 1986] G. McKoon & R. Ratcliff. Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 82-91, 1986.

[McKoon & Ratcliff 1989] G. McKoon & R. Ratcliff. Semantic associations and elaborative inference. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 326-338, 1989.

[Myers, Shinjo & Duffy 1987] J. L. Myers, M. Shinjo & S. A. Duffy. Degree of causal relatedness and memory. *Journal of Memory and Language*, 26, 453-465, 1987.

[Noordman & Vonk 1992] L. G.M. Noordman & W. Vonk. Reader's Knowledge and the Control of Inferences in Reading *Language and Cognitive Processes*, 7 (3/4) 373-391, 1992.

[O'Brien, Shank, Myers & Rayner 1988] E. J. O'Brien, D. M. Shank, J. L. Myers & K. Rayner. Elaborative inferences during reading: Do they occur online? J.E.P.: Learning, Memory, and Cognition, 14, 410-420, 1988.

[O'Brien, Duffy & Myers 1986] O'Brien, Duffy & Myers. Anaphoric inference during reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 12, 346-352, 1989.

[Sanford 1990] A. J. Sanford. On the nature of text driven inference. *Comprehension Processes and Reading*, D. A. Balota,

G. B. Flores D'Arcais and K. Rayner Eds., LEA, Hillsdale, NJ, 1990.

[Sanford & Garrod 1981] Sanford & Garrod. Understanding written language: Explorations in comprehension beyond the sentence. Chichester: John Wiley, 1981.

[Shastri & Ajjanagadde 1991] L. Shastri and V. Ajjanagadde. Rules and variables in neural nets. *Neural Computation*, 3(1):121–134, 1991.

[Shastri & Ajjanagadde 1993] L. Shastri and V. Ajjanagadde. From simple associations to systematic reasoning: A connectionist representation of rules, variables, and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences*, 16(4), 1993.

[Stevenson 1993] R. J. Stevenson. Language, Thought and Representation, John Wiley, Chichester, 1993.