

A Parser based on Connectionist Model

Hiroshi NAKAGAWA , Tatsunori MORI

Dept. Electronics and Computer Engineering, Yokohama National University
156 Tokiwadai Hodogaya-ku, Yokohama 240 Japan
net-mail-address a36646@cc.ccut.u-tokyo.junet

Abstract

This paper proposes a parser based fully upon the connectionist model(called "CM parser" hereafter). In order to realize the CM parser, we use Sigma-Pi-Units to implement a constraint of grammatical category order or word order, and a copy mechanism of sub-parse trees. Further more, we suppose there exist weak suppressive connection links between every pair of CM units. By these suppressive links, our CM parser explains why garden path sentences and/or deeply nested sentences are hard to recognize. Our CM parser also explains the preference principles for syntactically ambiguous sentences.

1. Introduction

In order to make clear a human parsing mechanism for natural language sentences, there remain some phenomena that are difficult to be explained by one integrated principle. These phenomena include cognitive difficulties to recognize garden path sentences or deeply nested sentences, and preference of structurally ambiguous sentences. All the parsing mechanisms proposed so far, for instances the top-down parsings /Pereira 1980/, the left corner parsing /Johnson-Laired 1983/, Marcus's parsing model/Marcus 1980/, Shieber's shift-reduce parser /Shieber 1983/, and so on, have not yet succeeded to explain all of these phenomena under one simple integrated principle. Note that all of them are based on symbol manipulation paradigm.

Recently a connectionist model (called CM hereafter) approach has been noticed in many area of cognitive science including natural language recognition. This approach has some advantages that the symbol manipulation approaches do not have. One advantage is that it is easy to use not only syntactic informations but also semantic and/or contextual informations in a uniform manner /Reilly 1984/. One fruitful result of this approach is the explanation about recognition of semantic garden path sentences like "The astronomer married the star" /Waltz 1985/. Another advantage is as follows. Since the connectionist model is a parallel system without any central controller, and an activation level of each unit and a connection strength between units may be presented as continuous values, it alludes much more flexible approaches than symbol manipulation approaches do. And we also expect it can simulate some aspects of human mental processing of sentence parsing.

This paper is concerned with the second advantage in parsing. The paper proposes a CM parser which can explain the above mentioned phenomena as preferences etc. in one integrated principle.

2. Parser based on connectionist model

Here we omit the technical details of the CM /McClelland&Rumelhart 1986/, but we must make clear that we stand for the so called "localist" view in which one symbol corresponds to one unit. Therefore in our CM parser, syntactical categories like noun phrase are represented by a unit in the CM, and a parse tree is represented as a network in which suitable syntactical categories being activated are connected. In order to realize a CM parser, we have to make clear the following two problems:

(1) How to express a word order or a syntactical

categories order appearing in phrase structure rules. For example, in a rule $S \rightarrow NP VP$, NP must precede VP.

(2) How to represent a case when a parse tree is generated by recursive phrase structure rules. Consider rules as follows: $S \rightarrow NP VP$, $NP \rightarrow NP \bar{S}$ and $\bar{S} \rightarrow Comp S$. The same pattern, in this case a pattern corresponding $S \rightarrow NP VP$, may appear more than once in a parse tree of one sentence. In order to represent this case, we need a copy mechanism of a partial parse tree pattern corresponding to the phrase structure rule in a connection network. Otherwise we have to prepare infinite number of copies of a partial parse tree pattern in advance. Of course this preparation is non-realistic not on computer hardware but on human wetware. In Fianty's CM parser mentioned in /McClelland&Kawamoto 1986/, the length of sentence is limited because of the above described preparation.

2.1 Phrase structure sub-network

Consider the next rule.

$$C \rightarrow A B \quad (3)$$

This rule has at least two meanings. One is that the category C consists of the category A and the category B. Another is that the category B follows the category A. This meaning is concerned directly with the problem (1). To represent a case that a word is coincident with some syntactic category, we modify (3) as follows.

$$C \rightarrow \text{word}$$

Since this rule is one variant of rule of type(3), we study about only rules of type (3) hereafter. We will explain about a sub-network that corresponds to the phrase structure rule (3).

We solve the problem (1) by introducing a trigger link that is presented as \rightarrow in figures. Namely "A \rightarrow B" expresses that B follows A. From the viewpoint of the CM, the meaning of this trigger link is that the unit for category B (called "B unit" hereafter) can be activated only when the unit for category A (called "A unit" hereafter) is fully activated. Due to the trigger link, the A unit must be activated chronologically faster than the B unit. The trigger link is realized by a Sigma-Pi-Unit /McClelland & Rumelhart 1986/ that includes a multiply operation. Figure 1 shows a concept of Sigma-Pi-Unit in the CM.

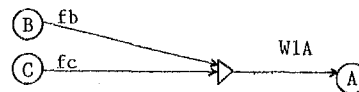


Figure 1. Sigma-Pi-Unit

In Figure 1, B and C are CM units. They send outputs whose values are fb and fc expressed as positive values , to the A unit. These values are corresponding to the B and C unit's activation levels respectively. WIA is a weight of link from B and C to A. The input to the A unit is as follows.

$$WIA * fb * fc$$

If the B unit's activation level:fb=0, then the C unit's activation level does not transmit to the A unit at all. In other words, the B(or C) unit's activation level is an on-off switch for activation transmission from the C(or B) unit to the A unit. Using Sigma-Pi-Units, a sub-network of phrase structure rule (3) is represented as shown in Figure 2. The weight $W_{A \rightarrow B}$ is very small in this case, but note that it depends on some semantic information.

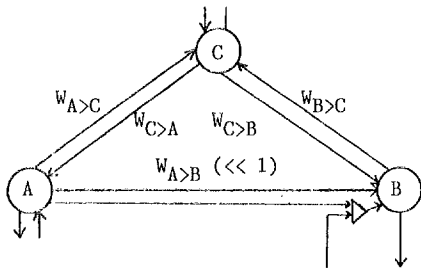


Figure 2. Sub-network of $C \rightarrow A B$

This network will be presented in a simpler form using a trigger link " $A \xrightarrow{t} B$ " hereafter as shown in Figure 3. A-, B-, and C-connectors' structures appeared in Figure 3 are explained in Section 2.3.

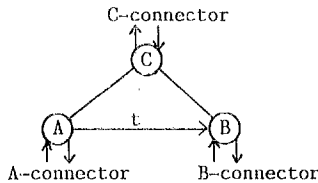


Figure 3. Simpler form of Figure 2's network

2.2 Copying sub-network

Our final goal is to make clear a mechanism of building a parse tree for a whole sentence by connecting sub-networks. For this purpose, the simplest method is preparing parse trees of all the possible sentence structures. In principle this method is not possible, because there are infinite number of possible sentence structures. Other method is preparing a number of copies of a sub-network for each phrase structure rule in advance. For example, ten sub-networks of $S \rightarrow NP VP$, ten sub-networks of $VP \rightarrow V NP$, and so on. When a parser reads a sentence, it selects some sub-networks from these prepared set of sub-networks, and connects them to make a parse tree of the input sentence. This method seems to work well and solves the above mentioned problem (2). Unfortunately this method has a serious deficiency as follows. From the view point of learning in the CM, all the weights of connection links of sub-networks are learned by parsing or recognizing a number of sentences. It is a plausible hypothesis that once a human becomes to be able to parse some structure of sentence, he/she ever can parse that structure since then. In order to explain this hypothesis, the above mentioned weights learning must be uniformly done for all copies of sub-networks of the same phrase structure rule. But this uniformly learning is too artificial for the human mental learning processes.

A solution avoiding these difficulties is as follows. There is only one central sub-network for one phrase structure rule, and all learning processes are done on it. In parsing, when a parser needs a sub-network of some rule, the parser makes copies of the sub-network and connects them into a suitable place of a parse tree yet to be constructed.

A sub-network copying mechanism is implemented as an application of the connection information distribution (CID) mechanism /McClelland 1986/. Figure 4 is a simple example of copying. The programmable sub-networks are implemented with the Sigma-Pi-Units. There are a lot of yet to be programmed programmable sub-networks, namely blank sub-networks. When the input comes in, the corresponding connection pattern of the central

network is copied to the programmable sub-networks via the connection activation system. In order to implement a copying mechanism of phrase structure rules in the form of $C \rightarrow A B$, we use three CID mechanisms. They are for bidirectional connections between the A unit and the C unit, between the B unit and the C unit, and between the A unit and the B unit respectively. We omit the further details because of the limited paper space.

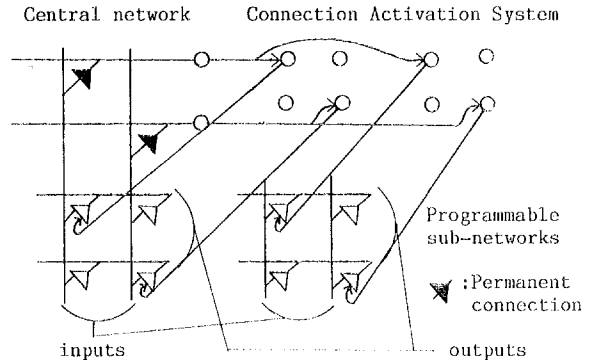


Figure 4. A simple connection information distribution (CID) mechanism

2.3 Connecting sub-networks

To generate a parse tree, we need a mechanism of generating connection links dynamically. Unfortunately the CM has not yet had this mechanism. Instead of this mechanism, we use a connector that changes connection dynamically by Sigma-Pi-Units. There are three kinds of connector, namely A-, B-, and C-connector as shown in Figure 3. We will explain

these connectors' functions in this section.

C-connector : If a C unit of a sub-network is activated, the C-connector sends requests for connection to A-connectors of blank sub-networks or B-connectors whose sub-network's B unit is the same syntactical category as the sender sub-network's C unit's syntactical category. More than one connections may be established by these requests, however, they suppress each other, and at last the connection from the most strongly activated B-unit wins. Even if a C unit is not so strongly activated, the C-connector sends these requests. Before a human has read a whole sentence, or even if he/she reads only few words, he/she predicts a complete or fairly large part of parse tree of possible sentence. This is why we adopt this low threshold strategy of requests sending.

A-connector : When an A-connector receives a request for connection from the other sub-network's C-connector, if the A-connector has not yet received any other requests for connecting, the A-connector makes a copy of sub-network whose A unit's syntactical category is the same as the syntactical category of C unit of the sender sub-network. By this copying, a parse tree grows in bottom-up manner.

B-connector : When a B-connector receives a request for connection from the other sub-network's C-connector, if the B unit's syntactical category is the same as the sender sub-network's C unit's syntactical category, a connection between the sender's C-connector and the receiver's B-connector is established. If more than one connections are established, they suppress each other. Finally the most strongly activated connection inhibits other connections. This suppressive or exclusive connections are expressed as $[X Y]$ shown in figures. In this expression, connections between X and Y are mutually suppressive or exclusive.

The above described connectors structure are shown in Figure 5,6 and 7 respectively.

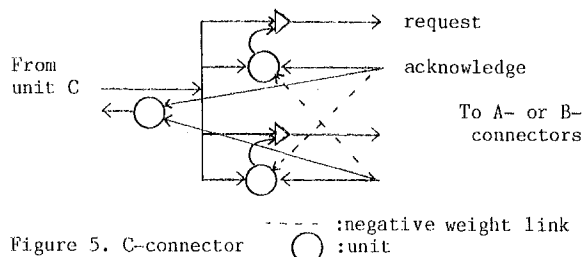


Figure 5. C-connector

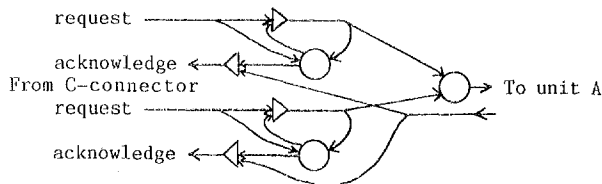


Figure 6. A-connector

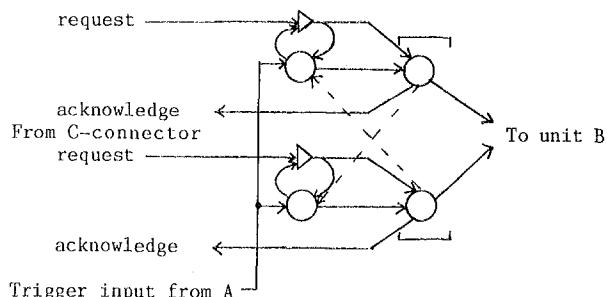


Figure 7. B-connector

2.4 Parsing on the CM parser

To summarize the above described CM parser, we sketch a parsing process of a sentence "I eat apples." Phrase structure rules used in this example are as follows. $S \rightarrow N VP$ and $VP \rightarrow V N$.

Parsing process

- (1) The CM parser reads "I", and a unit for category N is activated.
- (2) The C-connector of the N unit sends a request for connection to an A-connector of the currently usable blank sub-network.
- (3) When an A-connector receives the request, it makes a copy sub-network of $S \rightarrow N VP$. Since the N unit of the copied sub-network is fully activated, the trigger link from the N unit to the VP unit becomes active.
- (4) The CM parser reads "eats", and a unit for category V is activated, and a request for connection is sent from its C-connector to some A-connector.
- (5) When an A-connector receives this request, it makes a copy sub-network of $VP \rightarrow V N$. Not only the V unit but also the VP unit is activated. Of course the trigger link from the V unit to the N unit is activated.
- (6) The VP unit sends a request for connection via its C-connector. This request is received by the B-connector of the previously copied sub-network for the phrase structure rule $S \rightarrow N VP$, because this sub-network's B unit's category is VP, and the sender sub-network's C unit's category is also VP and triggered as you see at stage (3).
- (7) The CM parser reads "apples", and a unit for category N is activated, and a request for connection is sent from its C-connector.
- (8) This request is received by the B-connector of the copied sub-network at(5). This activates the C unit of this sub-network whose category is VP. This

activation causes that the B unit of the sub-network of $S \rightarrow N VP$. Finally, its C unit whose category is S becomes fully activated, namely the sentence is recognized and the parse tree is accomplished.

The result parse tree is shown in Figure 8. For compact expressions, the A- B- and C-connectors are omitted in the rest of the paper.

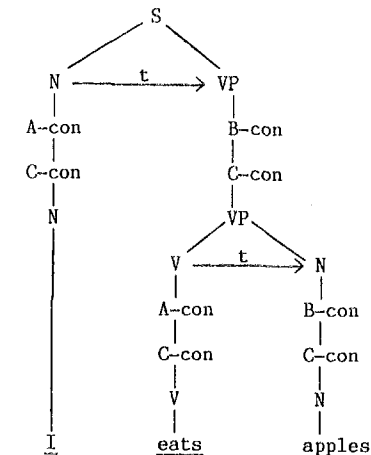


Figure 8. An example parse tree made by the CM parser

Intuitively, our CM parser is a parallel left corner parser. Speaking more precisely, owing to use a trigger link which predicts syntactic categories of the next incoming word, our CM parser is regarded as a parallel left corner parser with a continuous activation level for each generated nonterminal symbol representing some syntactic category.

3. Control on resource bounded condition

It is well known that a human memory system consists of at least two levels namely the short term memory and the long term memory respectively. A capacity of short term memory is limited to 7 ± 2 chunks. In the CM, an implementation of short term memory has not yet been cleared. But intuitively, the sum of all units' activation level is bounded. We implement this bound by the almost equivalent mechanism as follows. Namely there exist weak suppressive connection links between every pairs of units. Owing to this limitation, even if our CM parser is parallel one, it is impossible in parsing to maintain all possible candidate parse trees. Since our parser is based on the CM, the most promising parse tree is the most strongly activated one. Other parse trees are suppressed by the most promising one through the suppressive or the exclusive connections described in Section 2.3. In the rest of the paper, we propose explanations for control mechanisms of the CM parser especially about parsings of deeply nested sentences, garden path sentences and preferences of syntactically ambiguous sentences.

4. Recognition of deeply nested sentences

Our CM parser can explain why deeply nested sentences like "The man who the girl who the dog chased liked laughed" are hard to recognize for us human. Figure 9 shows a network being built just after the CM parser reads "The man who the girl who the dog chased". Here, since the NP_3 unit is strongly activated, the $VP_{2/NP}$ unit is strongly predicted and it is the right prediction. But since the NP_1 unit and the S unit are also activated, the VP_1 unit is also predicted. Therefore when the CM parser reads "liked", it is not very easy to select the $VP_{2/NP}$ unit definitely. As seen in this example, when the CM

parser reads a word at the deeply nested level, there may be a case that more than one units are strongly activated and predicted. If they have nearly the same activation level, it is not easy to select the right unit. This is one possible explanation why it is hard for us human to recognize deeply nested sentences, if the CM is a plausible model of the human mental process.

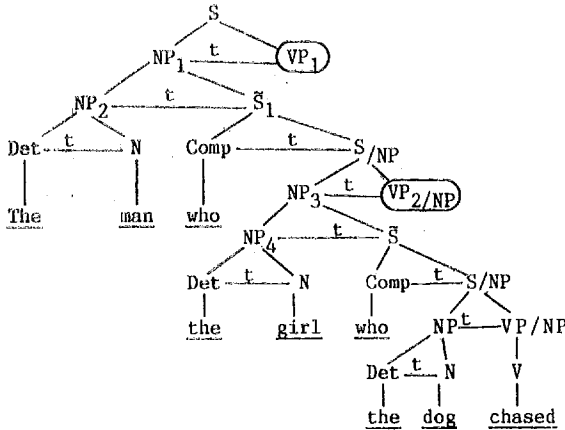


Figure 9. A parse tree (connected network) just after "The man who the girl who the dog chased"

5. Garden path sentences

If there are more than one possible syntactic structures for the input sentence, the CM parser makes more than one parse tree networks corresponding to them in a parsing process. If one of them is much more strongly activated than others, the parser easily selects it as the right network. But more than one networks are often activated to almost the same level. In the case, how to select one of them depends on many factors, for instance a contextual or a semantic information. There is a worse case as follows. Assume that a parser reads some words of the sentence, and there are more than one parse trees. One of them has the highest activation level than others at that time. But when the parser reads the next word, if the highest parse tree turns out to be syntactically impossible, some weakly activated parse tree is forced to be activated to the highest level suddenly. This forced sudden change of the activation level may cause us human a difficulty to recognize the sentence. This is an informal explanation for cognitive difficulty of recognizing garden path sentences.

In order to explain what parse tree is chosen, we have to recognize which exclusive connection plays the main role of preference between possible parse trees. Without loss of generality, it is sufficient to explain how one of two parse trees is chosen. In short, this choice point is such that an upper part of tree from this point is common to the both trees, and a part of trees that are below this choice point are different. Figure 10 shows a network generated for a garden path sentence "The cotton clothing is made of grows in Mississippi." The wrong parse tree including the S_a unit is preferred while our CM parser reads "The cotton clothing is made of", because in the phrase structure rule $\bar{S} \rightarrow S/NP$, the connection link from the S unit to the \bar{S} unit is weak, and "clothing" is NP. But when the CM parser reads "grows", the wrong parse tree including the S_a unit is rejected syntactically, and the right but weakly predicted VP_b unit must be connected to the VP unit for "grows". Maybe humans feel cognitive difficulty at that time. Note that although our CM parser should do a lot of works to parse a garden path sentence, namely the forced sudden change of

activation levels, finally it succeeds to parse the garden path sentence as well as human. It is a main difference of performance between our CM parser and Shieber's shift reduce parser.

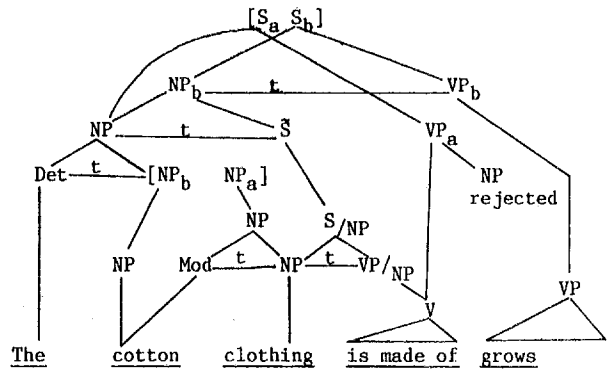


Figure 10. The parse tree network just after "The cotton clothing is made of grows"

6. Parsing Preference

If there are more than one possible syntactic structures for the input sentence after the entire sentence was input, one of them is preferred over others. In order to explain the parsing preferences, some syntactical preference principles such as Right Association, Minimal Attachment and so on, have been proposed so far in /Ford 1982/ et al. But there are some problems about these principles. The most important problem is which principle should be used in parsing the given sentence /Schubert 1984/. Since our parser is based on the CM, the parsing preferences are uniformly explained using each of the

activation level of the units being the components of parse tree for the given sentence. This preference mechanism with the activation levels is regarded as the minimal attachment principle for some cases and as the right association principle for other cases. In this section, we will show some examples about this matter.

The first example is about the sentence "John bought the book which I had selected for Mary." If we adopt the phrase structure rule $VP \rightarrow VP PP$, the result parse tree of this sentence generated by our CM parser is the one shown in Figure 11.

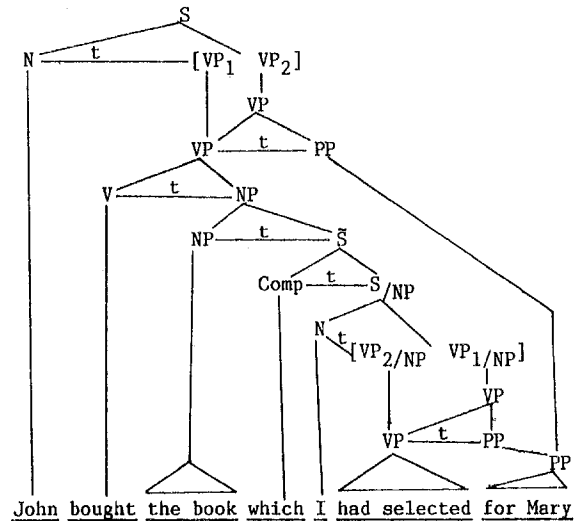


Figure 11. The example of parse trees of structural ambiguous sentence

There are two promising parse trees for this sentence as shown in Figure 11. If the tree including the VP₁ unit is preferred, the PP unit of "for Mary" is strongly connected to the VP₁/NP unit. If the tree including the VP₂ unit is preferred, the PP unit is strongly connected to the VP₂ unit. Now we examine the activation levels of these two unit. The VP₁/NP unit is activated directly by the V unit for "had selected". It is also indirectly activated and triggered by the N unit for "I". On the other hand, the VP₂ unit is indirectly activated by the V unit for "bought" and the NP unit for "the book which.." and so on. By this comparison, the VP₁/NP unit is known to be more strongly activated than the VP₂ unit. Therefore the PP unit for "for Mary" is more strongly connected to the VP₁/NP unit than the VP₂ unit, and the parse tree including the VP₁ unit is preferred. The result coincides with the right association principle that is likely used when humans parse this example sentence.

As you see in the example, many cases of the process of connecting to the most strongly activated unit are explained as the right association principle. But there are other cases in which the control mechanism are not so clear. Consider the next example.

"John carried the groceries for Mary."

Here we phrase structure rules of the Chomsky normal form. For instance, VP → VP PP, VP → V NP, and so on. The result parse trees are shown in Figure 12.

Notice that the native speakers of English show definite preferences for the parse tree including VP₂. Now we are required to explain a parsing control mechanism which causes this preference. If the PP unit for "for Mary" is connected to the NP₁ unit, the parse tree including the VP₁ unit is preferred. If the PP unit is connected to the VP₂ unit, the parse tree including the VP₂ unit is preferred. The NP₁ unit is activated directly by the NP unit for "the groceries". On the other hand, the VP₂ unit is

activated by both the NP unit for "the groceries" and the V unit for "carried" but indirectly. We can not determine which parse tree is preferred without further information for instance, the weight of every connection link. If the weight of the connection link from the VP unit to the VP₂ unit is very heavy, our parser prefer the parse tree including the VP₂ unit. From the viewpoint of phrase structure rules, by this connection link's heavy weight, we can regard the phrase structure rules VP → V NP and VP → VP PP as only one rule VP → V NP PP. Using this rule in parsing minimizes the resultant number of nodes. If we adopt the minimal attachment principle, the parse tree including the VP₂ unit is preferred. In short, the minimal attachment principle is explained in our parser's performance.

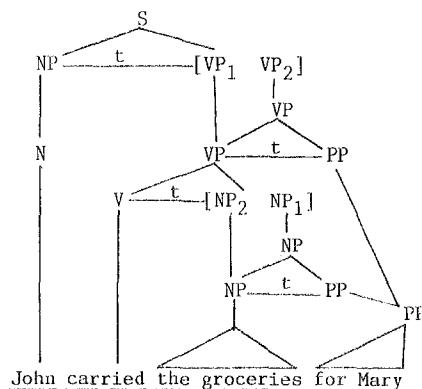


Figure 12. The parse trees for "John carried .." by VP → VP PP and VP → V NP etc.

As you know from these examples, the minimal attachment principle and the right association principle are integrated in our CM parser by determining the appropriate weights of connection links. This result is completely compatible the CM's principle that all informations are represented as connection link's weights.

7. Conclusions

We proposed a parser based fully on the CM. By introducing an upper bound for the sum of each unit's activation level into this CM parser, we can explain why garden path sentences and deeply nested sentences are hard to recognize. Our CM parser can integrate the minimal attachment principle and the right association principle into one principle that the most strongly activated unit is selected. Future work to be studied is to unify semantic and context informations into this CM parser.

Acknowledgement

We thank members of the special interest group of artificial intelligence so called "ATUFO", and Dr. K.Hashida at ETL. His elegant theory encouraged us to study about the work of this field. The research was supported by the Grant-in-Aid for Special Project Research of the ministry of education, science and culture, and the Inamori Foundation.

References

- Ford, M. Bresnan, J. & Kaplan, R. (1982), "A competence-based theory of syntactic closure", in Bresnan, J. (ed.), *The Mental Representation of Grammatical Relations*, MIT Press
- Jonson-Laired, P.N. (1983), "Mental Models" Cambridge University Press
- Marcus, M. (1980) "A Theory of Syntactic Recognition for Natural Language", MIT Press
- McClelland, J.L. (1986), "The Programmable Blackboard Model of Reading", in *Parallel Distributed Processing Vol.2*, The MIT Press
- McClelland, J.L. & Kawamoto, A.H. (1986), "Mechanisms of Sentence Processing: Assigning Roles to Constituents", in *Parallel Distributed Processing Vol. 2*, The MIT Press
- McClelland, J.L. & Rumelhart, D.E. et al. (1986), *Parallel Distributed Processing Vol.1 Vol.2*, The MIT Press
- Pereira, F.C.N. & Warren, D.H.D (1980) "Definite Clause Grammar for Language Analysis", *Artif. Intell.* 13
- Reilly, R.G. (1984) "A Connectionist Model of some Aspects of Anaphor Resolution" *COLING'84* pp.144-149
- Schubert, L.K. (1984) "On Parsing Preferences", *COLING'84* pp.247-250
- Shieber, S.M. (1983) "Sentence disambiguation by a shift-reduce parsing technique", *SRI International Technical Note 281*
- Waltz, D.L. & Pollack, J.B. (1985) "Massively Parallel Parsing: A Strongly Interactive Model of Natural Language Interpretation", *Cognitive Science* 9, pp.51-74